

```

1   TCGCGCCTTT CACCGGCACC TTGCGTCGGT CGCGCCGCGG GGCCTGCTCC
51  TGCCGCGCGC ACCCCCGGGG CTTCGGCTCC GGCACGGGTC GCGCCCAGCT
101 TTCCTGCACC TGAGGCCGCC GGCCAGCCGC CGCCATGGGT GCCTACCTCT
151 CCCAGCCCAA CACGGTGAAG TGCTCCGGGG ACGGGGTCGG CGCCCCGCGC
201 CTGCCGCTGC CCTACGGCTT CTCCGCCATG CAAGGCTGGC GCGTCTCCAT
251 GGAGGATGCT CACAACTGTA TTCCTGAGCT GGACAGTGAG ACAGCCATGT
301 TTTCTGTCTA CGATGGACAT GGAGGGGAGG AAGTTGCCTT GTACTGTGCC
351 AAATATCTTC CTGATATCAT CAAAGATCAG AAGGCCTACA AGGAAGGCAA
401 GCTACAGAAG GCTTTAGAAG ATGCCTTCTT GGCTATTGAC GCCAAATTGA
451 CCACTGAAGA AGTCATTAAA GAGCTGGCAC AGATTGCAGG GCGACCCACT
501 GAGGATGAAG ATGAAAAAGA AAAAGTAGCT GATGAAGATG ATGGGGACCA
551 CTTCTATAAG AGAAACAAGA ACCTGCCACC TGAGGAACAG ATGATTTTCA
601 CCCTTCCTGA CATCAAGGTG CTGACTCTCA CTGACGACCA TGAATTCATG
651 GTCATTGCCT GTGATGGCAT CTGGAATGTG ATGAGCAGCC AGGAAGTTGT
701 AGATTTTATT CAATCAAAGA TCAGCCAGCG TGATGAAAAT GGGGAGCTTC
751 GGTATTGTGC ATCCATTGTG GAAGAGCTGC TGGATCAGTG CCTGGCACCA
801 GACACTTCTG GGGATGGTAC AGGTGTGTGAC AACATGACCT GCATCATCAT
851 TTGCTTCAAG CCCCAGAAAC CAGCAGAGCT CCAGCCAGAG AGTGGCAAGC
901 GAAAACTAGA GGAGGTGCTC TCTACTGAGG GGGCTGAAGA AAATGGCAAC
951 AGCGACAAGA AGAAGAAGGC CAAGCGAGAC TAGCAGTCAT CCAGACCCCT
1001 GCCCACCTAG ACTGTTTTCT GAGCCCTCCG GACCTGAGAC TGAGTTTTGT
1051 CTTTTTCCTT TAGCCTTAGC AGTGGGTATG AGGTGTGCAG GGGGAGCTGG
1101 GTGGCTTCAC TCCGCCCATT CCAAAGAGGG CTCTCCCTCC AACTGCAGC
1151 CGGGAGCCTC TGCTGTCTCT CCCAGCCGCC TCTGCTCCTC GGGCTCATCA
1201 CCGGTTCTGT GCCTGTGCTC TGTGTGTGTT GAGGGAAGGA CTGGCGGTTC
1251 TGGTTTTTAC TCTGTGAACT TTATTTAAGG ACATTCTTTT TTATTGGCGG
1301 CTCCATGGCC CTCGGCCGCT TGCACCCGCT CTCTGTTGTA CACTTTCAAT
1351 CAACACTTTT TCAGACTAAA GGCCAAAACC TAATCGTTAA AAAAAAAAAA
1401 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAA (SEQ ID NO:1)

```

**FEATURES:**

5' UTR: 1-134  
Start: 135  
Stop: 981  
3' UTR: 984

**Homologous proteins:**

Top 10 BLAST Hits

|                    |  | Score | E     |
|--------------------|--|-------|-------|
| CRA 18000005093929 | /dataset=nraa /length=546 /altid=gi 4505999 ...  | 301   | 6e-81 |
| CRA 18000005157254 | /dataset=nraa /length=542 /altid=gi 6679793 ...  | 284   | 1e-75 |
| CRA 18000005062172 | /dataset=nraa /length=543 /altid=gi 3122591 ...  | 283   | 1e-75 |
| CRA 1000682325310  | /dataset=nraa /length=356 /altid=gi 7508306 /... | 161   | 1e-38 |
| CRA 18000004879328 | /dataset=nraa /length=348 /altid=gi 1352682 ...  | 161   | 1e-38 |
| CRA 18000004973158 | /dataset=nraa /length=370 /altid=gi 1171963 ...  | 154   | 9e-37 |
| CRA 18000005006718 | /dataset=nraa /length=414 /altid=gi 1171964 ...  | 154   | 1e-36 |
| CRA 18000004973159 | /dataset=nraa /length=414 /altid=gi 1076914 ...  | 152   | 3e-36 |
| CRA 89000000194666 | /dataset=nraa /length=352 /altid=gi 7291977 ...  | 151   | 8e-36 |
| CRA 18000004972853 | /dataset=nraa /length=464 /altid=gi 6320934 ...  | 145   | 4e-34 |
| CRA 89000000203611 | /dataset=nraa /length=662 /altid=gi 7302240 ...  | 141   | 9e-33 |
| CRA 18000004915558 | /dataset=nraa /length=468 /altid=gi 6319415 ...  | 140   | 2e-32 |
| CRA 18000005051471 | /dataset=nraa /length=468 /altid=gi 1622933 ...  | 140   | 2e-32 |

BLAST to dbEST:

|                      |                           |                |              |                  |  | Score | E   |
|----------------------|---------------------------|----------------|--------------|------------------|--|-------|-----|
| gb BF341923 BF341923 | 602016449F1               | NCI_CGAP_Brn64 | Homo sapiens | cD...            |  | 878   | 0.0 |
| gb AW370275 AW370275 | RC1-BT0254-131199-013-f12 | BT0254         | Homo sapi... |                  |  | 844   | 0.0 |
| gb BE305043 BE305043 | 601186706F1               | NIH_MGC_15     | Homo sapiens | cDNA c...        |  | 835   | 0.0 |
| gb BE249868 BE249868 | 600942906F1               | NIH_MGC_15     | Homo sapiens | cDNA c...        |  | 835   | 0.0 |
| gb AI417892 AI417892 | tg55b01.x1                | NCI_CGAP_Pr28  | Homo sapiens | cDNA...          |  | 823   | 0.0 |
| gb BE896031 BE896031 | 601438825F1               | NIH_MGC_72     | Homo sapiens | cDNA c...        |  | 815   | 0.0 |
| gb AI686098 AI686098 | tt92a01.x1                | NCI_CGAP_Pr28  | Homo sapiens | cDNA...          |  | 815   | 0.0 |
| gb AU143285 AU143285 | AU143285                  | Y79AA1         | Homo sapiens | cDNA clone Y7... |  | 813   | 0.0 |
| gb AU142966 AU142966 | AU142966                  | Y79AA1         | Homo sapiens | cDNA clone Y7... |  | 813   | 0.0 |
| gb BE903477 BE903477 | 601676724F1               | NIH_MGC_21     | Homo sapiens | cDNA c...        |  | 813   | 0.0 |
| gb BE891121 BE891121 | 601432236F1               | NIH_MGC_72     | Homo sapiens | cDNA c...        |  | 813   | 0.0 |
| gb BE886682 BE886682 | 601507935F1               | NIH_MGC_71     | Homo sapiens | cDNA c...        |  | 813   | 0.0 |
| gb BE797128 BE797128 | 601587118F1               | NIH_MGC_7      | Homo sapiens | cDNA cl...       |  | 813   | 0.0 |
| gb BE796239 BE796239 | 601591808F1               | NIH_MGC_7      | Homo sapiens | cDNA cl...       |  | 813   | 0.0 |
| gb BE793727 BE793727 | 601588924F1               | NIH_MGC_7      | Homo sapiens | cDNA cl...       |  | 813   | 0.0 |
| gb BE781616 BE781616 | 601467472F1               | NIH_MGC_67     | Homo sapiens | cDNA c...        |  | 813   | 0.0 |

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

Expression information from BLAST dbEST hits:

|                      |                           |
|----------------------|---------------------------|
| gb BF341923 BF341923 | Brain                     |
| gb AW370275 AW370275 | Breast                    |
| gb BE305043 BE305043 | Colon adenocarcinoma      |
| gb BE249868 BE249868 | Colon adenocarcinoma      |
| gb AI417892 AI417892 | Prostate                  |
| gb BE896031 BE896031 | Skin melanotic melanoma   |
| gb AI686098 AI686098 | Prostate                  |
| gb AU143285 AU143285 | Eye-retinoblastoma        |
| gb BE903477 BE903477 | Placenta Choriocarcinoma  |
| gb BE891121 BE891121 | Skin-melanotic melanoma   |
| gb BE886682 BE886682 | Uterus-leiomyosarcoma     |
| gb BE797128 BE797128 | Lung small cell carcinoma |
| gb BE796239 BE796239 | Lung small cell carcinoma |
| gb BE793727 BE793727 | Lung small cell carcinoma |
| gb BE781616 BE781616 | Eye-retinoblastoma        |

Expression information from PCR-based tissue screening panels:

Human leukocyte

1 MGAYLSQPNT VKCSGDGVGA PRLPLPYGES AMQGWRVSME DAHNCIPELD  
 51 SETAMFSVYD GHGGEVALY CAKYLPDIK DQKAYKEGKL QKALEDAFLA  
 101 IDAKLTTEEV IKELAQIAGR PTEDEDEKEK VADEDDGDHF YKRKNLPPPE  
 151 EQMISALPDI KVLTLTDDHE FMVIACDGIW NVMSSQEVVD FIQSKISQRD  
 201 ENGELRLLSS IVEELLDQCL APDTSGDGTG CDNMTCTIIC FKPRNTAELQ  
 251 PESGKRKLEE VLSTEGAEEN GNSDKKKKAK RD (SEQ ID NO:2)

# **FEATURES:**

## **Functional domains and key regions:**

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
 N-glycosylation site

233-236 NMTC

[2] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
 Protein kinase C phosphorylation site

Number of matches: 4

|   |         |     |
|---|---------|-----|
| 1 | 10-12   | TVK |
| 2 | 197-199 | SQR |
| 3 | 253-255 | SGK |
| 4 | 273-275 | SDK |

[3] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
 Casein kinase II phosphorylation site

Number of matches: 10

|    |         |      |
|----|---------|------|
| 1  | 38-41   | SMED |
| 2  | 57-60   | SVYD |
| 3  | 106-109 | TEEE |
| 4  | 122-125 | TEDE |
| 5  | 164-167 | TLTD |
| 6  | 184-187 | SSQE |
| 7  | 197-200 | SQRD |
| 8  | 210-213 | SIVE |
| 9  | 224-227 | TSGD |
| 10 | 229-232 | TGCD |

[4] PDOC00008 PS00008 MYRISTYL  
 N-myristoylation site

Number of matches: 2

|   |         |        |
|---|---------|--------|
| 1 | 2-7     | GAYLSQ |
| 2 | 266-271 | GAEENG |

[5] PDOC00009 PS00009 AMIDATION  
 Amidation site

253-256 SGKR

[6] PDOC00792 PS01032 PP2C  
 Protein phosphatase 2C signature

55-63 MFSVYDGHG

**BLAST Alignment to Top Hit:**

>CRA|18000005093929 /dataset=nraa /length=546 /altid=gi|4505999  
/def=ref|NP\_002698.1| protein phosphatase 1G (formerly  
2C), magnesium-dependent, gamma isoform; protein  
phosphatase 1G (formerly 2C),; protein phosphatase 2,  
catalytic subunit, gamma isoform [Homo sapiens]  
/org=Homo sapiens /taxon=9606  
Length = 546

Score = 301 bits (763), Expect = 6e-81  
Identities = 146/146 (100%), Positives = 146/146 (100%)

Query: 137 GDHFKRNKNLPPEEQMISALPDIKVLTLTDDHEFMVIACDGIWNVMSQEVVDFIQSKI 196  
GDHFKRNKNLPPEEQMISALPDIKVLTLTDDHEFMVIACDGIWNVMSQEVVDFIQSKI  
Sbjct: 401 GDHFKRNKNLPPEEQMISALPDIKVLTLTDDHEFMVIACDGIWNVMSQEVVDFIQSKI 460

Query: 197 SQRDENGELRLLSSIVEELLQCLAPDTSGDGTGCDNMTCTIICFKPRNTAELQPESGKR 256  
SQRDENGELRLLSSIVEELLQCLAPDTSGDGTGCDNMTCTIICFKPRNTAELQPESGKR  
Sbjct: 461 SQRDENGELRLLSSIVEELLQCLAPDTSGDGTGCDNMTCTIICFKPRNTAELQPESGKR 520

Query: 257 KLEEVLSTEGAEENGNSDKKKKAKRD 282  
KLEEVLSTEGAEENGNSDKKKKAKRD  
Sbjct: 521 KLEEVLSTEGAEENGNSDKKKKAKRD 546 (SEQ ID NO:4)

Score = 284 bits (718), Expect = 1e-75  
Identities = 137/139 (98%), Positives = 138/139 (98%)

Query: 1 MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPELDSETAMFSVYD 60  
MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPELDSETAMFSVYD  
Sbjct: 1 MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPELDSETAMFSVYD 60

Query: 61 GHGGEEVALYCAKYPDIKDKQKAYKEGKLQKALEDAFLAIDAKLTTEEVIKELAQIAGR 120  
GHGGEEVALYCAKYPDIKDKQKAYKEGKLQKALEDAFLAIDAKLTTEEVIKELAQIAGR  
Sbjct: 61 GHGGEEVALYCAKYPDIKDKQKAYKEGKLQKALEDAFLAIDAKLTTEEVIKELAQIAGR 120

Query: 121 PTEDEDEKEKVADEDDGDH 139  
PTEDEDEKEKVADEDD D+  
Sbjct: 121 PTEDEDEKEKVADEDDVDN 139 (SEQ ID NO:5)

>CRA|18000005157254 /dataset=nraa /length=542 /altid=gi|6679793  
/def=ref|NP\_032040.1| fibroblast growth factor inducible  
13 [Mus musculus] /org=Mus musculus /taxon=10090  
Length = 542

Score = 284 bits (718), Expect = 1e-75  
Identities = 139/146 (95%), Positives = 141/146 (96%), Gaps = 1/146 (0%)

Query: 137 GDHFKRNKNLPPEEQMISALPDIKVLTLTDDHEFMVIACDGIWNVMSQEVVDFIQSKI 196  
GDHFKRNKNLPP+EQMISALPDIKVLTLTDDHEFMVIACDGIWNVMSQEVVDFIQSKI  
Sbjct: 398 GDHFKRNKNLPPQEQMISALPDIKVLTLTDDHEFMVIACDGIWNVMSQEVVDFIQSKI 457

Query: 197 SQRDENGELRLLSSIVEELLQCLAPDTSGDGTGCDNMTCTIICFKPRNTAELQPESGKR 256  
SQRDENGELRLLSSIVEELLQCLAPDTSGDGTGCDNMTCTIICFKPRNT ELQ ESGKR  
Sbjct: 458 SQRDENGELRLLSSIVEELLQCLAPDTSGDGTGCDNMTCTIICFKPRNTVELQAESGKR 517

Query: 257 KLEEVNSTEGAEENGNSDKKKKAKRD 282  
KLEE LSTEGAE+ GNSD KKKAKRD  
Sbjct: 518 KLEEALSTEGAEDTGNSD-KKKAKRD 542 (SEQ ID NO:6)

Score = 279 bits (706), Expect = 3e-74  
Identities = 133/139 (95%), Positives = 138/139 (98%)

Query: 1 MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPELDSETAMFSVYD 60  
MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPELD+ETAMFSVYD  
Sbjct: 1 MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPELDNETAMFSVYD 60

Query: 61 GHGGEEVALYCAKYLPDIKDQKAYKEGKLQKALEDAFLAIDAKLTTEEVIKELAQIAGR 120  
GHGGEEVALYCAKYLPDIKDQKAYKEGKLQKAL+DAFLAIDAKLTTEEVIKELAQIAGR  
Sbjct: 61 GHGGEEVALYCAKYLPDIKDQKAYKEGKLQKALQDAFLAIDAKLTTEEVIKELAQIAGR 120

Query: 121 PTEDEDEKEKVADEDDGDH 139  
PTEDED+K+KVADEDD D+  
Sbjct: 121 PTEDEDDKDKVADEDDVDN 139 (SEQ ID NO:7)

Hammer search results (Pfam):

| Model   | Description            | Score | E-value | N |
|---------|------------------------|-------|---------|---|
| PF00481 | Protein phosphatase 2C | 176.2 | 4.2e-49 | 2 |
| PF01722 | BolA-like protein      | 3.7   | 9.5     | 1 |

Parsed for domains:

| Model   | Domain | seq-f | seq-t  | hmm-f | hmm-t  | score | E-value |
|---------|--------|-------|--------|-------|--------|-------|---------|
| PF00481 | 1/2    | 25    | 102 .. | 1     | 95 [.] | 74.1  | 4.6e-20 |
| PF01722 | 1/1    | 99    | 110 .. | 76    | 87 .]  | 3.7   | 9.5     |
| PF00481 | 2/2    | 137   | 228 .. | 197   | 301 .] | 102.2 | 4.5e-28 |

|      |             |             |             |            |            |
|------|-------------|-------------|-------------|------------|------------|
| 1    | AAAGAATCTT  | TTTTTTTTTT  | TTGAGACGGA  | GTTGCTCTGT | CACCCAGGGT |
| 51   | GGAGTGCAGT  | GGCGCCATCT  | TGGTTCACCTG | CAACCTCCGC | CTCTGGGGTT |
| 101  | CAAGCTATTC  | GCCTGCCTTA  | GCCTCCCAAG  | TAGCTGGGAT | TACAGGAGCG |
| 151  | CACCACTACG  | CCTGGCTAAT  | TTTTGTATTT  | TTAGTAGAGA | CGGGTTTCAC |
| 201  | ATGTTGGCCA  | GGCTGGTCTC  | GAACCTCTGG  | CCTCAAGTGA | TCCACCACCC |
| 251  | CCCCTTGGCC  | TCCCAAAGTG  | CTGGGATTAC  | AAGTGTGAGC | CACTGTGCCC |
| 301  | GGCTGAAAAG  | AATCAATTTT  | GTCATAGTTT  | GGAGAATTTT | TCCTTTTCTC |
| 351  | TCCATCCCTT  | GAATGCAATT  | TATTACCAAA  | TCTGTCTTAT | TTGTTATTGT |
| 401  | CTAATTTGTC  | CTTTCATCTG  | GATTCCCATT  | GCCACCCTGC | GTGGTACCAC |
| 451  | CTTACTCCCA  | GCTCTTCTCA  | TCTCCTGCTT  | AGAGTAAGAG | CTCTCTAACT |
| 501  | AGTAGCAGTG  | CCCCAGGCCA  | GGCGCGGTGG  | CTCACGCCTG | TAATCCCAGC |
| 551  | ACTTTGGGAG  | GCTGAGGCGG  | GTAGATCACG  | AGGTCAGGAG | TTCGAGACCA |
| 601  | GCCTGGCCAA  | CATGGTGAAA  | CCCCGTCTCT  | ACTAAAAATA | CAAAAATTAG |
| 651  | CCAGGCGTGG  | AATCCTAGCT  | ACTCGGGAGG  | CTGAGGCAGG | AGAATCGCTT |
| 701  | GAACCTGGGA  | GGCGGAGGTT  | GCAGTGAGCC  | AAGATAGCGC | CACTGCACTG |
| 751  | CAGCCTGGGC  | AACAAGAGCG  | AAACTCTGTC  | TTAAAAAAA  | AACAATAGTA |
| 801  | GGCGGGGTGC  | GGTGGCTCAC  | GCCTGTAATC  | CCAACACTTT | GGGAGGCCGA |
| 851  | GGCGGGCGGA  | TCACGAGGTC  | AGGAGATGGA  | GACCATCCTG | GCTAACACGG |
| 901  | TGAAACCTCG  | TCTCTACTAA  | AAATACAAAA  | AATTAGCCAG | GCGTGGTGGC |
| 951  | GGGCGCCTGT  | AGTCCCAGCT  | ACTCGGGAGG  | CTGAGGCAGG | AGAATGGCGT |
| 1001 | GAACCCGGGA  | GGCGGAGCTT  | GTAGCCTGGG  | CGACAGAGCG | AGACTCTGTC |
| 1051 | TCAAAAACAA  | ACAAAAACA   | ACAACACAAC  | AGTGCCCCAG | ACTCTCTCCC |
| 1101 | TCCAATGTAC  | ACTGCATACA  | AAGACTAGAC  | AAACGATGCC | AAAGGTTACA |
| 1151 | CCAGGACGAC  | AATGAAGTCC  | AAGTCACTCA  | TCTTGGCGCT | TCTTTTTATC |
| 1201 | AAGCTAGCTA  | GTTAATATTA  | CCACTTACAA  | TCATTCTCCG | AGTCCAGCCA |
| 1251 | GCTTCCTCAG  | AAATCCTCCC  | CAAATGCAGT  | TCACATTCTT | CCCTCACTCC |
| 1301 | TCCAAACCTT  | ATAACATTAT  | TCCCTTTTCC  | TGTGGCATAA | TGCAATCCAG |
| 1351 | AGGCATCCTC  | TCTGAGAGCC  | TACATCCTTT  | CAATCCTCCA | AGAAACAGCT |
| 1401 | CCTCCTCCCA  | TCCTTGAGCT  | TTTCCCACCC  | AGAATAGGCT | GTACCAAACA |
| 1451 | TTTCTACATT  | GTATTAATC   | GAATTATATC  | ACAGATCGAG | GTCTTCGCTT |
| 1501 | TCAGAAAAGA  | CTCACATTCT  | TCTCATATAG  | GCCTCGCATG | GCCTGGCCAA |
| 1551 | GTAATTCCCTA | AGCTCGGAAC  | AAATACTGGT  | CAACTTTAAT | TGAACCAAAT |
| 1601 | CGGGCGGGGT  | TTGCGGAGTC  | TGAGAGTGCA  | ACGTTGGGGA | GAGGGGGATG |
| 1651 | AAAACACTGG  | AGGACGGGCT  | GAAAGCGTCG  | AGTCCGACAC | AAAAGAGGCG |
| 1701 | TCAGACAAAA  | CGCCAAGAGG  | CTGGGGACTG  | GGAACGAAGG | AAGAGGTTCT |
| 1751 | GCCAGAGGCG  | ACCTGCCACC  | TGCGCGAGGA  | AGCGGAGTAG | GACGGCGGCC |
| 1801 | GTTGGTGGGC  | GTGGTTCGCG  | TAGTCTCGCG  | GGAGCGGCCG | TTGGGCGGGC |
| 1851 | CGTTGTCCCC  | TGCGGGCGGG  | GCGAGTTGCT  | AAGGAAATGA | CTGCCCCGAG |
| 1901 | CGCCTGGCCC  | CGCCGCGCAG  | GCCGGGCGGG  | GTCTGGAGCG | GCGCCGTTTC |
| 1951 | CGCTTCCGCT  | CCCTCACAGC  | TCCCGTCCCG  | TTACCGCCTC | CTGGCCGGCC |
| 2001 | TCGCGCCTTT  | CACCGGCACC  | TTGCGTCCGT  | CGCGCCGCGG | GGCCTGCTCC |
| 2051 | TGCCGCGCGC  | ACCCCCGGGG  | CTTCGGCTCC  | GGCACGGGTC | GCGCCCAGCT |
| 2101 | TTCTTGACAC  | TGAGCCGCCG  | GCCAGCCGCC  | GCCATGGGTG | CCTACCTCTC |
| 2151 | CCAGCCCAAC  | ACGGTGAAGT  | GCTCCGGGGA  | CGGGGTCCGC | GCCCCGCGCC |
| 2201 | TGCCGCTGCC  | CTACGGCTTC  | TCCGCCATGC  | AAGGCTGGCG | CGTCTCCATG |
| 2251 | GAGGTGAGGA  | GGCAGGGGCC  | CATAGGCTGG  | CCGCTGCGGG | GCGGGAATCT |
| 2301 | GACGGAGAAA  | GAGAGCGGGG  | GATGGGGTCC  | TCCCCTGGGA | AGGGTCCCAA |
| 2351 | TTGGGAGCCT  | GCGGCCGCGC  | CGGCCGTTTG  | CGGGGCGACA | GAGACCGCGG |
| 2401 | GGTCAGGGCC  | GAGAGGGAGC  | TCCCGATGCT  | TGGGGACCGT | GTGCCGGTGG |
| 2451 | CCAGTGGCGG  | CGAGGGCTCG  | GCCATGTGGG  | AAGAGGCACC | TTCCGCCAC  |
| 2501 | TGACCGCCCT  | CTCCCCGAGC  | TTTGGCGCCA  | TCCTCTCGTG | CCAACCTAGC |
| 2551 | CCTCCAGGCT  | CATCAACGGT  | GTGAGGTTTA  | GTGTGGGAGT | AAAGACGCAA |
| 2601 | ATAGGGGCTA  | TTTATTCATT  | TGTTTTTCAA  | GGGAGAGGTG | AATAGATGTG |
| 2651 | AATAACTTTT  | TAAATTTTAA  | TATTTAAAT   | ATCTGATGTG | GGAAGCCTCT |
| 2701 | TTTGGCTAGG  | AGTTTGACAG  | TGAAAGGAAC  | CCCGGGCAGA | GTCTGTTTCA |
| 2751 | CATTTTGGTT  | GCCTGGCCTT  | GGGCTCTTGC  | ATGTTAATTT | CAGAGGCTGG |
| 2801 | ACCCGACCTC  | CAGGAGTTGT  | CACTCATTTG  | CACTCTTTTC | AGGGCCTTTT |
| 2851 | ACTAACTTCG  | GAAAACTTGA  | ATTATGTCAG  | TCCCTAGGTT | TTCTTTTATA |
| 2901 | TTATGCTTCG  | TTTTCTTTCT  | CTCTTAGGAT  | TTCTCTAAAA | CTTAATCAGT |
| 2951 | AATTCTCATT  | TGCCTGTAAT  | TGTAGTTTAT  | AGTTTCTTTG | ATTGTAAGTC |
| 3001 | ATTTTCATCA  | TTTTTTTTCAT | CACAACCTAC  | CTACAAAGGG | CTTTTCTAGA |
| 3051 | AAATTTTACT  | CTGGACAAAA  | GGGGAAAAGA  | AAAATATTGG | GGGGAAAGTA |
| 3101 | GTAGTATTAG  | GTAAACTTG   | ATGTGAAACT  | ACAAAAGAGA | AGAGGGAAAA |

FIGURE 3, page 1 of 17





|      |            |             |            |            |             |
|------|------------|-------------|------------|------------|-------------|
| 6301 | TCAGCCATTT | CTGACTTTTG  | ATCTGTGTTT | TCTCTTTGGT | TTCTCTTTTT  |
| 6351 | TTTTTTCTTT | TTTTTTTTTT  | GAGACGGAGT | GTCGCTCTGT | TGCCCAGGCT  |
| 6401 | GGAGTGCAGT | GGCATGATCT  | TGGCTCACTG | CAACCTCTGC | CTCCTGAGTT  |
| 6451 | CAAGCGATTC | TCCTGCCTCA  | GCTTCCCAG  | TAGCTGGGAT | TACAGGCGCC  |
| 6501 | TGCCACCATG | CCCGGCTAAT  | TTTTTGTATT | TTCAATAGAG | ACAGGGTTTC  |
| 6551 | ACTATGTTGG | CCGGGTTGGT  | CTCCAACCTC | TGACCTCTTA | ATCCGCCCCG  |
| 6601 | CTCGGCCTCC | CAAAGTGCTG  | GCATTACAGG | CGTGAGCCAC | CATGCCTGGC  |
| 6651 | CCATCAGTTG | ATGTAGTCTT  | AAGGGGACAA | GAGTACATTT | AATATTTGGT  |
| 6701 | TGTGAGGTTT | TCTGGAAGTG  | ACAAAACCTG | TTTCTATGGA | GAGTTAGGAA  |
| 6751 | TTTTTTTTTT | TTTTGAAACG  | GAGTCTCGCA | TTGTCACCCG | GGCTGGAGTG  |
| 6801 | TAGTGGCTTG | ATCTCGGCTC  | ACTGAAACTT | CCGCCTCTTG | GGTTCAAGTG  |
| 6851 | ATTCTCCTGC | CTCAGCCTTC  | CAAATAGCTG | GGATTACAGG | AGTCTGCCAC  |
| 6901 | CAGGCCAGCT | AATTTTTTTT  | TGTATTTTTT | GTGGAGACAG | GATTTCACTA  |
| 6951 | TGTTGGCCAG | GCTGGTCTCA  | AGACTCCTGA | CGTTGTGATC | CACCTGCCTC  |
| 7001 | GGCCTCCCAA | AGTGCTGGGA  | TTACAGATGT | GAGCTACCGT | GCCCCGCCAG  |
| 7051 | GAATTTTTTG | TGCTATAAAT  | CATATTTTCC | TTTTATTAAA | GGCAGTGTCA  |
| 7101 | ATATCTATAG | TATAATTTTG  | AGGAGGCTGG | CTATTTATTG | CTGTGTAGAA  |
| 7151 | GCTGGCTTAT | TAGTGGTCAA  | GGGGTCATCT | AGAATTGACT | ATAAAGATAG  |
| 7201 | TATTGAGCAG | AAAATTCTTA  | AAATACCTGC | ATATTAGTTT | CAGTCATTAA  |
| 7251 | ATTAATGGAA | AAAATATAAA  | AAGAAATATC | ACAAGTATGC | TATGGGGTTC  |
| 7301 | TACCTTAGGG | CATTGAAGGT  | TGAAAAACAT | TTATTTTCTT | ATCTTCAGAT  |
| 7351 | TAGCATCTCA | TATCAATAGC  | CAATAGCTTA | AAGCGCTTTT | TACTTACTAA  |
| 7401 | ACCAGGTCAG | AATTTCTCTC  | TCTCTCTCTT | TTTTTTTTTT | AGACAGAGTC  |
| 7451 | TCTGTACACC | AGTCTAGAGT  | GCAGTGGTGT | GATCTCGGCT | CACCTGGCCTT |
| 7501 | TCTGTTTAAG | TGATTCTCCG  | GCCTCAGCCT | CCTGAGTAGC | TGGGATTACA  |
| 7551 | GGCATGTGCC | ACGACACTTG  | GCTTATTTTT | TGTATTTTTT | GTAGAGATGG  |
| 7601 | GGTTTCTCTG | TGTTGGTCTG  | GCAGGTCTTG | AACTCCTGAC | CTCAGGTGAT  |
| 7651 | CTGCCCCCCT | CCGCCTCCCA  | AAGTGCTGGG | ATTACAGACG | TGAGCTACTG  |
| 7701 | CGCCTGGCCA | GAATTTCTTT  | GTCTAGAATG | TGGTTAGCAA | CTTTTATAAA  |
| 7751 | AACGCATTAT | TTGCATTTGA  | TTAGCATGCA | GTACCCATTC | ACAGTTCAAA  |
| 7801 | GCTAGTATAG | AATTATATCA  | CATGTATGCC | CATGAGCATG | GAGAAACTAT  |
| 7851 | TTTCTTTTTT | TTTTTTTTAAG | TTGGAGTTTT | GCTCTTGTTG | CCCAGGCTGG  |
| 7901 | AGTGCAATGG | TGCCATCTCG  | GCTCACTGCA | GCTTCTGCCT | CCTGGGTTCT  |
| 7951 | AGCAATTTGC | CTACCCAGC   | CTCCCAAGTA | GCTGGGATTA | CAGGCACTCG  |
| 8001 | CCACCATGCC | CAGCTAACTT  | TTTTGTATTT | TTAGTAGAGA | AGGGGTTTTT  |
| 8051 | CCATGCTGGC | GAGTCTGGTC  | TTGAACCTCT | GGCCTCAAGT | GATCTGCCCC  |
| 8101 | CCTCAGCCTC | CCAAAGTGCT  | GGAATTACAG | GCATGAGCCA | CTGTGCCCCG  |
| 8151 | CCTTTTTATT | TTTTAAATTA  | TTTATGTATT | TATTTTGAGA | CAGGATCTCA  |
| 8201 | CTCTTGCCCA | TGCTTGAGTG  | GTAAGGAGTA | TGGGATTTCT | TGTGCCGGTT  |
| 8251 | CCTTCACATC | CTCACTACAC  | TTATCTGCCT | TTACAGTGGC | TCGATCATAG  |
| 8301 | TTCACTGCAT | AGCCTTCTGG  | GCTCAAGGGG | TCTTCCAGCC | TCAGCCTAAT  |
| 8351 | ATAGGCACAT | GCCACCATGA  | CTGGCTAATT | TTTTTTTTTT | AAGTTTTTTT  |
| 8401 | TTGTAGAGAT | AGGGCCTTGC  | AGTGTTGCC  | AGGCTGAGGA | ATTTTATTTA  |
| 8451 | TGTTTATTTT | ATTTATTTAT  | TTATTTATTT | ATTTATTTAT | TTATTTATTT  |
| 8501 | ATTGAGATGG | AGTCTTACAC  | TGTCACCCAG | GCTGGAGTGC | AGTTGCGCGG  |
| 8551 | TCTCAGCTAC | TGCAAGCTCC  | GCCTCCCGGG | TTCATGCCAT | TCTCCTGCCT  |
| 8601 | CAGCCTCCCC | CGTAGCTGGA  | CTACAGGCGC | CCGCCACCAC | GCCTGGCTAA  |
| 8651 | TTTTTTGTGT | GTGTTTTTAG  | TAGAGGCGGG | GTTTCACCAT | GTTAGCCAGG  |
| 8701 | ATGGTCTCGA | TTCTTGACCT  | CGTGATCCAC | CCACATCGGC | CTCCCAAAGT  |
| 8751 | GCTGGGATTA | CAGGTGTGAG  | CCACCATGCC | TGGTCTAGAA | ATTATTTTAT  |
| 8801 | ATTTTATACC | ATTGCCTTAT  | AAGTTCTCAA | GCAACTGGAA | AATACAATCA  |
| 8851 | GAACGTATTC | CTCAAGATTT  | CAAGGATATT | TTACACAAAG | TTCTATTGTC  |
| 8901 | TGATTCCTTA | GCAGTTGTTA  | CTACTGTTTC | CCTAACCTCT | AATCTTCTAT  |
| 8951 | TGGGTTATTA | GTCTTAGAAT  | TGAATTTTGA | GAGGTAAGGG | CTTGAATTTG  |
| 9001 | AACATAGAAA | TTTATACAGG  | TCTGATCAGT | AGTTCTTGAC | ATTGTATTAT  |
| 9051 | CTGGAAACAA | ATCTTTAGAA  | CTGAGCTTAA | GATGTTTAAT | GACATTTTGT  |
| 9101 | AGACAGAGTA | TGATTTTCTG  | GATGTTGTTT | TTGTTTCTTT | CTAGATCTAG  |
| 9151 | TTCAAGAGAT | AAGTATATCA  | ACTTTTTTTT | CCTTTTTGAC | CCAATGCTAG  |
| 9201 | CAGAAAAACA | ACACCTTTTA  | ATCATATTTA | GTATTTGAAA | ATGTGTATAC  |
| 9251 | AGGTTCCCTT | TTATTTTATT  | TATTTCTTAC | AGGTTCCCTT | TTAATCAGCT  |
| 9301 | TTATTGAGAT | AGAGTTCATA  | TACTGTATGG | TTCATACCAC | ATATGGTTCA  |
| 9351 | TATACCATAC | AGTATATGAA  | CTCACTTTAA | GAGTATAATT | CAGTGGGTTT  |
| 9401 | AAGGGTATAA | TTCATTCATT  | TTAAGGGTAT | AATTCAGTGG | CTTTTAGTAT  |

FIGURE 3, page 3 of 17

|       |             |            |            |            |             |
|-------|-------------|------------|------------|------------|-------------|
| 9451  | ATTTTCTTTT  | TTTTCTTCTT | TTTTTCTTTT | TTCGAGACAG | GGTCTTGCTC  |
| 9501  | TGTTACCCAG  | GCTGGAGTGC | AGTGACGCAG | CCTCAGCTCA | CTGCAACCTC  |
| 9551  | CACCTCCCAG  | GTTCAAGCGA | ATTCTCTTGC | CTCAGCATCC | TGAGTAGCTG  |
| 9601  | GAACACAGG   | CTCACGCCAC | CACACCCAGC | TAATTTTTAT | ATTTTCAGTA  |
| 9651  | GAAACAGGGT  | TTCACCATGT | TGCCCAGGCT | GGTCTCGAAC | TCCTGGCCTC  |
| 9701  | AAGTGATCTG  | CCTGCCTCAG | CCTCCCAAAG | TGCTGGGATT | ACAGGCATGA  |
| 9751  | GCCACCGTGC  | CTGCCTGTTT | TGTAGTGTAT | TCAAACAGAG | TTGTACAACT  |
| 9801  | GTCACCACAA  | TCAGTTTTAG | AACCCCAAAA | AGAAACCCTG | TACTCTTTAC  |
| 9851  | CAGTCACTCC  | CTATCTTCCG | TCCACTAACC | CCTGGCATCC | ACTAATTTAC  |
| 9901  | ATGACCTCTA  | TGAATTTGCC | TATTCTGAAC | ATTTTATAAA | TGGAATTCTA  |
| 9951  | AATACACTAC  | CTTTTATATC | TGGATGCTTT | TACTAAGCAT | GTGTATTTTT  |
| 10001 | GAAATTGACT  | TAAAGCTTG  | TTGGCCCCTG | GAAGAGTAAA | TTACTCTCCA  |
| 10051 | CCCCAAGTAT  | TCCCTCTACC | CCTCAGCTTT | GCCTGTAAGT | TTCTTTTTAA  |
| 10101 | AAAAATCACA  | CATACATTGT | TGTAGTAGAT | TTAAGAATAA | GTATTTTGCT  |
| 10151 | GACCCAAGGT  | TCTTTTGCTT | CTTTCTAGAT | CAGTGCCTTG | CAGGTTTTAT  |
| 10201 | TTTACAGAGC  | TTAATAGAAT | CAGAAATCTC | TTTAAAACTC | CAGTCTCATA  |
| 10251 | TCCAGTTATC  | ACTCACCATC | TCTGTGTTTG | CAGCAATAGC | CAGGCCTGGC  |
| 10301 | CCAGAGGGAC  | TTGATCTCCA | CTTTTGCTTT | TTAGACTTTT | CTGTGGCTTT  |
| 10351 | TACCACCTGC  | TGTGTATCCT | TGACCTCATA | CTGCTGGACT | CCTTTGGATG  |
| 10401 | GATACCAGCA  | GGATGGTTCA | GGCTCCAGTG | GGCACTTTTT | AAAATTCTCT  |
| 10451 | CCTTCTGTTC  | AGATAGACAG | AGCTCAGGCA | GATCACCAAG | TCTGTTGCCT  |
| 10501 | GTGTAACCAG  | GAAAGAGATG | CTAGTTTTCT | TTAGGCACT  | CCCATTTGTT  |
| 10551 | TCTGTTGGAA  | CCTTCCTCAC | TTAGTTGATG | GAAGGGAAGC | AAAAGACCCA  |
| 10601 | GAACCTCATC  | TCAAATTAAT | GACTTAACAA | TTCTTGAATT | TTCTCTTATC  |
| 10651 | TCCTAGTTAA  | CTCTTTTCT  | TATCTCCAGG | AGGTCAGTTT | TAATTATTGT  |
| 10701 | TGTTTATTTCA | TTTATTTTCA | TGGAGACAGG | GTCTTACTGT | GTTGTCCAGG  |
| 10751 | CTGGCCTTGA  | ACCCCTGGCC | TCAAGGAATC | TTTTCACCTC | AGCCTCCCAA  |
| 10801 | AGTGCTGGAA  | TTACAGGCAT | GAGTCACCAC | ACCCAGCCTG | ATATTTTTCA  |
| 10851 | GTTGATGTAT  | CATAGTTGTG | CCTAAGCATA | ATTTTTTAAT | TTTAATTTTT  |
| 10901 | TATTTTTGGG  | GACAGGGTCT | CCCTCTGTCT | CCCAGGCTGG | AGTGCAGTGA  |
| 10951 | TGCGATCTCA  | GCTCACTGCA | ACCTCCACCT | TCTGGGTAA  | AGCGATTCTC  |
| 11001 | CTGCCTCAGC  | CTCCCGAGTA | GCTGGGACTA | CAGGTACCCA | CCATCACACC  |
| 11051 | CGGCTAAATT  | TTTTGTGTGT | ATTTTAGTAG | AGACGAGATT | TCGCTGTGTT  |
| 11101 | GCCCAGGGCG  | GTCTCGAACT | CCTGAGCTCA | GGCAATCCGC | CTGCCTCGGC  |
| 11151 | CTCCCAAAGT  | GCTGGGATTA | CAGGCATGAA | CCACCACGCC | CGGCCAAGCG  |
| 11201 | TAATATTTTT  | AAGGGTCATC | AATGTTGTGT | CATGAATCAA | TCAGTGTTTC  |
| 11251 | GTTCTTTTTT  | ATGGTTGAAT | AATATTCCAT | GGTATGGATT | TGTCACATTT  |
| 11301 | TGTTTATCCA  | TTCATTAGTT | GATAGACATT | TTGGATTTC  | ACTTTTTTTT  |
| 11351 | TTTTTTTTTT  | GCTATTATAA | ATAGTGATAC | TATGTACAAA | TTTTTGTGTG  |
| 11401 | GAAATATGTC  | CTCATATCTC | TTGGTTATAT | ACCAAAGAGT | GGAAGTGCTG  |
| 11451 | GGTCATATGG  | TAACACGTG  | TTTAACATTT | TGAGAAACTG | CTAAACTGTT  |
| 11501 | TTCCAAAGTT  | GCTGTACCGT | CGTACATTCC | TGCCAGCAAT | ATATGAGGAT  |
| 11551 | GCCAGTTCCT  | TCACATGTTC | ACTACACTTA | TCCACCTTTT | TTATAATAAC  |
| 11601 | TAATGGTGGG  | TGTGAGATGG | TATCTCATTG | TAGTTTTGAT | TTGTATTTCT  |
| 11651 | CTGATGGCTA  | AATGGCTAAT | GATGTTTGAA | CTTTTGTGTT | GAGACAGAAT  |
| 11701 | CTCACTCTGT  | CCAGATTCAA | GCGATTCTCC | TGCCTCAGCC | TCCCTAGCAG  |
| 11751 | CTGGGATTAC  | AGGCACATGC | CACCACACCC | AGCTAATTTT | TTGTATTTCT  |
| 11801 | AGTAGAGACA  | GGGCATTACC | ATGTTGTTCA | GGCTGGTGTC | GAACCTCTGA  |
| 11851 | CCTCAAAGGA  | TCCGCCTCCC | TGGGCCTCCC | AAAGTGCTGG | ATTACAGGCT  |
| 11901 | AGAGCCACCA  | TGCCAGGCCT | TATGTTTGAA | CATCTTTTAT | GTGCTTATTG  |
| 11951 | GACATTTGTG  | TATCTTCTTT | GGAGAAATGT | CTGTTCAAAG | TCTTTGTCCA  |
| 12001 | TTTTTAATTG  | GATTGTCTTT | TTGTCTTTTG | ATGTGTAAGA | GTTCTTTATG  |
| 12051 | TGTTTTGGAT  | ACAAGTTTGT | TAGATATATG | ATTTGCAAAT | CTTTTCTCCA  |
| 12101 | ATTTTGTGG   | ACTTTTGCTT | TCTTTTTTTG | TTTTGTGTTT | GTTGTTGTTG  |
| 12151 | TTGTTGTTGT  | TGTTTTGGTC | GGGGGACAGT | CTTGCTCTGA | CCACCCAGGC  |
| 12201 | TGGAATGGAG  | TGGCGCGATC | TTGGTTCACT | GCAACCTCTG | CCTCCTGAGT  |
| 12251 | TCAAGCTATC  | CTGCTTCAGC | CTCCCGAGTA | GCTGGGACCC | AGGTGTGTGC  |
| 12301 | CACCACTCCC  | AGCTAATTTT | TTATTTTTAG | TAGAGACCGG | GTTTCACCAT  |
| 12351 | GTTGGCCAGG  | CTGGTCTTGA | ACTACTGACC | TCAGGTGATC | TGCCCTGCCTC |
| 12401 | AGCCTCCCAA  | AGTGCTGGGA | TTACAGTCAT | GAGCCACTAC | ACCCTGATTC  |
| 12451 | TTTTTGCTGG  | CTTCTTTTCT | TTTTTTTTCT | TTTTTTTTTT | GAGACGGAGT  |
| 12501 | CTCGCTCTGT  | TGCCAGGCTG | GAGTGCAGTG | GCATGATCTC | GGCTCACTGC  |
| 12551 | AACCTCTGCC  | TCCCGGGTTC | AAGCCATTCT | CCTGCCTCAG | CCTCCCGAGT  |

FIGURE 3, page 4 of 17

|       |             |             |            |            |             |
|-------|-------------|-------------|------------|------------|-------------|
| 12601 | AGCTAGGACT  | ATAGGCACAT  | GCCACCATGC | GCAGCTAATT | TTTGTATTTT  |
| 12651 | TAGTAGAGAC  | GGGGTTTCAC  | CATGTTGGCC | AGGATAGTCT | CGATCTCTTG  |
| 12701 | ACCTCGTGAT  | CCGTCGCGCT  | GGGCCTCCCA | AAGTGCTAGG | ATTACAGACG  |
| 12751 | TGAGCCACCA  | CACTCAGCCT  | CTTTTTGCTT | TCTTGATGGT | GTCTTTTGAA  |
| 12801 | ACAAAAGTTT  | TTACTTTTGA  | TAAAGTCCAA | TTTGTCTATT | TTGTTTGTTT  |
| 12851 | GTTTTTGTTA  | AGAAGCTTTG  | CCTAACCCAA | AGTCACGAGA | ATTTTCTCTT  |
| 12901 | AGGTTTTCTT  | CTAAGAGTTT  | TATAGTTTTA | GCTGTTTCTG | TGATCCATTT  |
| 12951 | TGAGTGAATT  | TTTGTGAATG  | GTATGAGGGA | GTGATCCAAC | TTCATTCTTT  |
| 13001 | TGTGTGTGGA  | TATCAAGTTG  | TCCCAGCACT | ATTTGTTTAA | ACCACTGTTT  |
| 13051 | TTTTCCCCCA  | TTGAATTATC  | TTGGCATCAT | TGTCAGAGAT | AAATTGACCG  |
| 13101 | TAAATGTGAG  | GGTTTTATTT  | CTGAACCTCT | AAGTCCATTT | CATTGGTCTA  |
| 13151 | CATGTCCCTA  | TGCCAGTAAT  | ACACTATCTT | GGTACTGTGA | GCTTTTTTAGT |
| 13201 | ACGTTTTGAA  | ATGTTTTTAA  | AATTTGTTTT | TCATCTAAAT | TTTAGGATTA  |
| 13251 | ATTTGTCAAT  | TTCTGCACAA  | AAGGCACCTG | GGTTTCTATA | GGGGTTATGC  |
| 13301 | AGAATCTGTA  | GATCAACTGG  | GGGAGTATTA | CAGGCATGAG | CCACCGTGCC  |
| 13351 | TGGCTGACTG  | AGTTTTTTCAT | AGATGTACTC | TATCAGGTTT | AGGAAGTTCC  |
| 13401 | CTTTTATPCC  | TAGGTTGTTG  | AGTCTATTTT | ATATTACTTT | TTTAGAGACA  |
| 13451 | GTCTTGCTCT  | GTCCCTCAGG  | CTGTAGCACA | GTGGCTCAAT | CATAGCTCAC  |
| 13501 | TGCAGCCTTG  | AACTCCTAGG  | TTCAAGAGAT | CCGCCTGCCT | CAGCCTTCTT  |
| 13551 | AGTAGCTGGG  | ATTACATGCA  | TGCACCACCA | TACTGGGCTA | ATTTTTTAAA  |
| 13601 | ATTTTTTATA  | GAGACAGGGT  | CTTATTACTA | TGTTGCCCAG | ACTGGCATTG  |
| 13651 | AGTCTTTTAA  | TCATTAATGA  | GCACTGAATT | TTGTCAAGTG | CCTTTATAAT  |
| 13701 | ACCTATTGTG  | ATGATCATAG  | GGTTTTGTTC | TTTAGTCTAC | CGATACGCTA  |
| 13751 | TATTGCATTA  | AGTGATTTTT  | TTGAATGTTA | AACCAACCTT | GCATTTTTTTT |
| 13801 | GGTGTATAAG  | TCTTATTTGA  | TCAATGTGTA | TTATCCTTTT | ATATGGTGCT  |
| 13851 | GGATTTAGTT  | TGCTACTATT  | TTGTTGAGGA | TTTTTGTGTC | TATATTCATA  |
| 13901 | AGAGATATTG  | GTCTGTAGTT  | TCTTGTGATG | TCTTTGTCTG | GTTTTAGAAT  |
| 13951 | CAGGGTAATG  | CTGGCCTCAT  | AGAATGAATT | GGGAAGTGTT | GTCTTTTCTA  |
| 14001 | TGTGATGGGA  | GAGTTTGTGA  | ATCATTGGTA | TTAATTTTTT | TGTAAATGTT  |
| 14051 | TGGTAGAATT  | CACAAATAAA  | GGCATCTGAG | CCTGGGCTCT | TCTTTGTGGG  |
| 14101 | AAGTTTTTGG  | CTTTTTTTTT  | CTTTAAAAAT | TTTCATTGTG | GCTGGGCATG  |
| 14151 | GTGGCTCACG  | CCTGTAATCC  | CAGCACTTTG | GGGGGCCAAG | ACGGGTGAAT  |
| 14201 | CACCTGAGGN  | NNNNNNNNNN  | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN  |
| 14251 | NNNNNNNNNN  | NNNNNNNNNN  | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN  |
| 14301 | NNNNNNNNNN  | NNNNNNNNNN  | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN  |
| 14351 | NNNNNNNNNN  | NNNNNNNNNN  | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN  |
| 14401 | NNNNNNNNNN  | NNNNNNNNNN  | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN  |
| 14451 | NNNNNNNNNN  | NNNNNNNNNN  | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNGAG  |
| 14501 | TGCAATGATT  | GCAGTCTTGG  | CTCACTGCAA | CCTCTGCCAC | CTGGGCTCAA  |
| 14551 | GAGATTCTCC  | TGCCTCAGCC  | TCCTGAGTAG | CTGGGATTAC | AGGCATGCGC  |
| 14601 | CACCATGCCC  | AGTTAATTTT  | TGTATTTTTA | GTAGAGACGG | GGATTCTCCA  |
| 14651 | TGTTGACCAG  | GCTGGTCTCT  | AACTCCTCAC | CTCAAGTGAT | CCGCCCCGCT  |
| 14701 | CGGCCTCCCA  | AAGTGCTGGG  | ATTACAGGCA | TGAGCCACCA | CGCCCCGCTT  |
| 14751 | TAAAAATTTT  | TTTAATGTAC  | AGTTGAGTAG | TATTTAATAC | ATTCACATTG  |
| 14801 | TTGTGTACCC  | AGTTTCCAGA  | ACTCTTCATC | CTACAGAACT | GAAACTCCAT  |
| 14851 | ACCCATTAAA  | TGAGTCCCCA  | TTCTCTTTCC | CCCAGCTCAT | GGCAAACAGC  |
| 14901 | ATTCTATTTT  | CAGTCTCTAT  | GAATTTGATT | AGTTTAGATA | CTTCATACTG  |
| 14951 | TAAGTGGAAT  | CATATGGTAT  | TTGTCTTTTA | GTGACTGCCT | AATTTAAAAA  |
| 15001 | AAATTTTTTT  | GAGACGGAGT  | CCTGCTCTGT | CGCCCAGGCT | GGAGTGCAGT  |
| 15051 | GGCACCATCT  | CTGCTCACTG  | CAACCTCCAC | CTCCCAGGTT | CAAGTGATTG  |
| 15101 | TCCTGCCCTCG | GCCTCCCACG  | TAGCTGGGAT | TACAGGTGCT | CGCCACAACA  |
| 15151 | CCCGGCTAAT  | TTTTGTATTT  | TTAGGTAGAG | ACTGGGTTTC | ACCATGTTGG  |
| 15201 | CCAGGCTGGT  | CTCGAACTCC  | TGACCTCAAA | TTATCCACCT | GCCTTGGCCT  |
| 15251 | CCCAAAGTGC  | TGGGATTACA  | GGCGTGAGCC | ACTGTGCCCA | GCCTCCATGT  |
| 15301 | TGTTTTTCAC  | AACACCTGTA  | TCATTTACAT | TTCCACCAAC | AGTACACAAG  |
| 15351 | AATTTTCAGT  | TCTCCACATC  | CTTGCTAGCA | GTTGTTATTA | TCTGTTTTTT  |
| 15401 | TTTAATGGTT  | TCTTTTTTCC  | TTTTTCTTTT | TTTTTTTTTT | TGAGACGGTC  |
| 15451 | TTATTCTTGC  | TCATGCTGGA  | GTGCAGTGGT | GCAATGTGAT | AGCTCACTGC  |
| 15501 | AGCCTCAACC  | TCTGGGCTCA  | AGCAGTCCTG | CCACCTCAGC | CTCCACATAG  |
| 15551 | GTGGGACTGC  | AGGTGTGCAC  | CACCACTTGT | GGCTAATTTA | AAAAATTTTT  |
| 15601 | TCGTAGAGAC  | AGAGTCTCAC  | AGTGTTACCC | AGGCTGGTCT | TGAACTTCTG  |
| 15651 | AGCACAAGTG  | ATCCTCCCAC  | CTCAGCCTCC | CAAAATAATG | AGATTAGAGA  |
| 15701 | CATGAGCCAA  | CATGCCCAAC  | CAGTTTTGTT | TGTTTTGTTT | GTTTTGTGTT  |

FIGURE 3, page 5 of 17

|       |            |             |            |            |            |
|-------|------------|-------------|------------|------------|------------|
| 15751 | TTTGAGACAG | AGTCTCACTC  | TATTGCCCAG | GCTGGAGTGC | AACGGCATGA |
| 15801 | TCTTAGCTCA | CTGCAACCTC  | CGCCTCCCAG | GTTCAAGTGA | TTCTCATGCC |
| 15851 | CCAGCCTCCT | GAGAAGCTGG  | GATTACAGGT | GTACCACCAC | ACCCAGTTAT |
| 15901 | TTTTGTATTT | TTAGTAGACA  | TGGGGTTTTG | CCATGTTGGC | CAGGCTGGTC |
| 15951 | CCGAACTCCT | GACCTCAAGT  | GATCTGCTCC | CCTCAGCCTC | CCAAGGTGCT |
| 16001 | AGAATTAAGT | TTTTCTTTCT  | TTCTTTCTTT | CTTTTTTTTT | TTTTTTTTGA |
| 16051 | GACAGAGTCT | CACTCTGTCA  | CCCAGGCAGG | AGTGCAATGG | CACGGTCTTG |
| 16101 | GCTCATTGTA | ACCTCTGCCT  | CCCAGATTCA | AGTAGTGATT | CTCCTGTCTC |
| 16151 | AGCCTCCCAA | GTAGCTGGGA  | TTACAGGCAT | GCACCACCAC | GCCCAGCTAA |
| 16201 | TTTTTTGTAT | TTTGTAGTAG  | AACGGGGTTT | CACCATGTTG | GTCAGGCTGA |
| 16251 | TCTCAAACCT | CTGACCCCAA  | GTGATCCACC | CGCCTTGGCC | TCCCAAAGTG |
| 16301 | TTGGGATTAC | AGGCGTGAGC  | CACTGTGCCT | GGTTTTATTT | TTATTATTAT |
| 16351 | TATTTTTAAT | AGTTCCTATT  | CTAATGGGTA | TGAGGTAGTG | AGGTGGGTGG |
| 16401 | TTGTGGTGTT | TTTATGAATG  | TTTAATTGGA | AATGGGTGGC | CATTGTGTGC |
| 16451 | AGGAAAAACC | TCCTAAATTG  | TGTCAAACCT | CTGGAAAATG | AAATATCATT |
| 16501 | CCAGTTGCAA | GAATATCTTT  | TTTTTTTTTT | TTTTTTTTTA | AGACAGAGTC |
| 16551 | TCACTCTGTC | ACCAGGCGGG  | AGTGCAGTGG | CACGATCTCG | GCTCACTGCA |
| 16601 | ACCTCCGCCT | CCTGGTCCGC  | CTCCCGGGTT | CAAGTGATTC | TTCTGCCTTA |
| 16651 | GCCTCCCAAG | TAGCTGGGAC  | TACAGGCGCG | TGCCACCACT | CCTGGCTAAT |
| 16701 | TTTTTTGTAT | TTTGTAGTAG  | GATGGGGTTT | CACCATGTTG | GCCAGGATGG |
| 16751 | TCTCAATCTC | TTGACCTAGA  | GATCCGCCTG | CCTCGGCCTC | CCAAAGTGGT |
| 16801 | GGGATTACAG | GGGGGTCACC  | GTGCCCAGCC | ACAAGAAGAT | CTTGAGCATG |
| 16851 | TGAATGATCA | GAAATGATTT  | AGCCTATGTA | GGCACTAGGC | CAGGTAGTGA |
| 16901 | AATTCAGGGA | AAATAATTCA  | GATGCTTCTG | AGCTATCACT | TATGAACATA |
| 16951 | GAAACAGCTT | AAAGCCATTA  | TAGTGTGTTT | CCTGAAGATG | AAAGCATATG |
| 17001 | GTAAGATGAA | ATAGTGATTA  | TTTTTTAAAA | ATTACTACTC | CAGAAAGGAA |
| 17051 | AAGTTTACTA | ATTTTTATTA  | CTAAAGTTTA | CTGTTGGTGG | GTGCGGTGGC |
| 17101 | TCACACCTGT | AATCCCAGCA  | CTTTGGGAGG | CCGAGGCAGG | CGGATCACCT |
| 17151 | GAGGTCAGGA | GTTTCGAGAC  | AGCCTGACCA | ATATGGTGAA | ACCCCATCTC |
| 17201 | TGCTAAAAAT | AAAAAATTAG  | GCCGGGCGCG | GTGGCTCATG | CCTGTAATCC |
| 17251 | CAGCACTTTG | GGAGGCCGAG  | GCAGGTGGAT | CACGAGGTCA | GGAGATTGAG |
| 17301 | ACCATCCTGG | CTAACACGGT  | GAAACCCCGT | CTCTGCTAAA | AATACAAAAA |
| 17351 | TCAGCCGGGC | GTCTTGCCAG  | GCACCTGTAG | TCCCAGGTAC | TCAGGAGTTT |
| 17401 | TGAGACGGGA | GAATGGCGTG  | AACCCGGAAG | GCGGAGCTTG | AAGTGAGCCG |
| 17451 | AGATTGCGCC | GCTTCAGTCC  | AGCCTGGACG | ACAGAGTGAG | ACTCTGTCTC |
| 17501 | TAAAAAAATA | AATAAATAAA  | AATAAAAAAT | TAGCTGGGTG | TGGTGGCACG |
| 17551 | CACCTGTACT | CCCAGCTACT  | CGGGAGGCTG | AGGCAGGAGA | ATTGCTTGAA |
| 17601 | CCCGGGAGAT | GGAGGTTGCG  | GGGAGCCAAG | ATTGCGCCAC | TGCACTCCAG |
| 17651 | CCTGGCGACA | GAGTGAGACT  | CTTTCTCAGA | AAAAAATATG | ATAATTAAAA |
| 17701 | GTTGAGACGT | TCTTCGCCGA  | GAGTGGTCCG | GGTTTCCTGC | TTCAACAGTG |
| 17751 | CTTGACGGA  | ACCCGGCGCT  | CGTCCTGCAC | CCCGGCCGGC | CGCCCATAGC |
| 17801 | CAGCCCTCCG | TCACCTCTTC  | ACCATGCCCT | CGGACTGCCC | CAAGGCCCCC |
| 17851 | GCCGCAGCTC | CAGCGCCGCG  | TAGCCACCAC | TGCCGCTGCC | GCCGCCTCTC |
| 17901 | CTTAGTCGCC | GGCATGACGA  | CCGCGTCTAC | CTCGCAGGTG | CGCCAGAACT |
| 17951 | ACCACCAAGA | CTCAGAGGCC  | GCCATCAACC | GCCAGATCAA | CCTGGAGCTC |
| 18001 | TACGCCTCCT | CCATTTACCT  | GTGCGTGGCT | TACTACTTTG | ACAGCGATGA |
| 18051 | TGTGGCTTTG | AAGAACTTTG  | CCAAATACTT | TCTTCACCAA | TCTCATGAGG |
| 18101 | AGAGGGAACA | TGCTGAGAAA  | TTGATGAAGC | TGTAGAACCA | ACGAGGTGGC |
| 18151 | CGAATCTTCC | TTCAGGATAT  | CAAGAAACCA | GACTGTGCGG | GGAGAATGCG |
| 18201 | ATGGGAGAGC | GGGCTGAATG  | CGATGGATTA | CATTTGGAAA | AAATTGTGCA |
| 18251 | TTTTGCATTA | CATTTGGAAA  | AAAATGTGAA | TCAGTCACTA | CTGGAAGTGC |
| 18301 | ACAAACTGGC | CACTGACAAA  | AATGACCCCC | ATTTGTGTGA | CTTCATTGAG |
| 18351 | ACACATTACC | TGAATGAACA  | AGTGAAGGCC | ATCAAAGAAT | TGGGTGACCA |
| 18401 | CGTGACCAAC | ATGCACGAGA  | TGGGAGCGCC | CGAATCTGGC | GTGGCAGAAT |
| 18451 | ACCTCTTTGA | CAAGCACACC  | CTGGGAGACA | GTGATAATGA | AAGCTAAGCC |
| 18501 | TCAGGCTAAT | TTCCCCATAG  | CCATAGGGTG | ACTTACCTTG | TCACCAAGGC |
| 18551 | AGCGCATGTA | TGTTGGGGTT  | TCCTTTACCT | TTTCTATAAG | TTGTTCCAAG |
| 18601 | ACACCCACTT | AAGTTCTTTG  | ATTTGTACCA | TTCTTTCAAA | TAAATAAATT |
| 18651 | TGGTACCCTC | CCCCCCCCCA  | AAAAAAAAT  | GTAAGTGGG  | CTGGCGTAGT |
| 18701 | GGCTCATGCC | TAAATCTCAG  | CACCTTGGGA | GGCTGAGGCG | GGAGGATCAC |
| 18751 | CTGAGGTCGG | GAGTTTGAGA  | CCAGCCTGGG | CAACATGGTG | AAACCCCGTG |
| 18801 | TCTACTAAAA | ATATAAAAAAC | TAGCCAGTCA | TGGTGGCACA | CACCTGTAAT |
| 18851 | CCCGGCTACT | TGGGAGGCTG  | AGGCATGAGA | ATCACTTGAA | CCTGGGCTGC |

FIGURE 3, page 6 of 17

|       |            |            |            |            |            |
|-------|------------|------------|------------|------------|------------|
| 18901 | GGAGGTTGTA | GTAAGCTGAG | ATCATGCCAC | TGTACTCCAG | CCTGGGTGAC |
| 18951 | AGGGAGACAT | TCTCTCTCTC | AAAAAAAAAA | AAAAAACAAA | AAAAACAAAA |
| 19001 | CAAACCAACA | AAACAAAGTA | ATCCAGGAAC | AACAACATGA | TGAAGGACTG |
| 19051 | CATGCAGGAC | TCAGTGATGG | ATGGTGGAAG | ACAGCCAGGA | AGTTAAGCAT |
| 19101 | GACTCTGGTA | TTAAGTGTTG | TCTGGGAGAG | TTAAGATTCC | ATTTACAGAA |
| 19151 | ATAAGACCTG | TAGGGGAAGC | TCTTGATTTT | TTTTTTTTTT | CAGACTGCTG |
| 19201 | ATTTCTTGAT | TACATGTGTT | AAGTTTGAGG | TATAGAGAGA | AAGAACATCC |
| 19251 | TGGCCGGGTG | CAGTGGCTCA | CACCCGTAAT | CCCAGCACTT | TGGGAGGCCA |
| 19301 | AGGTGGGCAG | ATCACGAGGT | CCAGGAGATC | GAGACCATCC | TGGCCAACAT |
| 19351 | GGTGAAACCC | CGTCTCTACT | AAAAATACAA | AAATTAGCTG | GGCGTGGCGG |
| 19401 | CGCGTGCCTG | TTATCCCAGC | TACTCAGGAG | GCTGAGGCAG | GAGAATTGCT |
| 19451 | TGAACCCGAG | AGGCAGAGGT | TGTGATGAGC | CGAGATCGCG | CCACTGCACT |
| 19501 | CCACCCTGGC | AACAGAGCTA | GACTCTGTCT | CAAAAAAAAA | AAAAAGAAAG |
| 19551 | AAAAAAGAA  | CATCCTGTAG | AAACAGGCAG | TCAGAGGTAT | AGAACTACAC |
| 19601 | AGAATCCAAG | AGATCTTTCA | AGAAAAGTGA | CATGCAGCAA | GAGAAACTAT |
| 19651 | CAAGGGGGTA | AACAACCTAT | AGAATGGGAG | AAAATATTCA | CAAAGTATAC |
| 19701 | ATCCAACAAA | AGTCTAATAT | CCAAAATCTA | TAAGGAACCT | AACAAGCAAA |
| 19751 | AAGCAAATAA | CCCCCTTAAA | AAGTGGGCAA | AGGACATGAA | CAGATACTTC |
| 19801 | TCAAAAGACG | TACATGTGGC | CCACAAACAT | GAAAAAACGC | CCATTTCTAA |
| 19851 | TCATCAGACA | AATGCAAATT | AGAACCACAA | AGAGATACCA | TCTCACACCA |
| 19901 | GTCAGAACAG | CTTTTGTTAA | AATGTCAAAA | AATGAGAAAC | GTTGGTGAGG |
| 19951 | CTGCAGAGGA | AAGCAGACAC | TTGTACACTG | TTGGTGAAGG | TGTAAATTAA |
| 20001 | TTTAGCGTAG | GCACAGTCAG | TTTGGAGATT | TCTCAGAGAA | CTAAGAGTGG |
| 20051 | AACTACCATT | AGACCCAGCA | ATCTGATGGC | TGGATATACG | GCCAAAGGAA |
| 20101 | AATAAATCAT | TCTGCCAAAA | GAACATATGT | ACCTGTATGT | TCATTGTGGC |
| 20151 | ACCATTCACA | ATAGCAAAAA | CATTGAATCA | ACTCATGTGC | CCATCAGTGG |
| 20201 | CGGACTAGAA | AAGAAAAGAA | AATATGGTAC | ATAGCCATCA | TGGAATACTA |
| 20251 | TGCACCCATT | AAAAATAATG | AAATAATGTC | TTTGCAACAA | CATGAATGTA |
| 20301 | GCTGGAGGGC | ATTATCCTAA | GCAAACCTAA | ACAGAAACAG | AAAACCAAAT |
| 20351 | ACTGCGTGTT | CTCACGCAGT | GAGAGTGGGA | GCTGAACATC | AAGTACACAT |
| 20401 | GGATGTAAAG | ATGGCAACAA | TAGACATGGG | TCTACTAGAG | GTGGTGGTGC |
| 20451 | GGCAGGGTGG | GGGTGGGGGT | TGTGTGGCAG | AGGAACAGCT | GAAAAACTAC |
| 20501 | CTATTTGATA | CTATACCCAG | CACCTGGGAA | ACGGGTTTCA | TCATACCCCA |
| 20551 | AACCTCAGCA | TCACACAGTA | TACCTTTCTA | ACAAACTTAC | ACATGTATTC |
| 20601 | TGTGATTCTA | AAATAAACAT | TGAAAAATAA | AAAAAAACT  | GACATGGTTT |
| 20651 | GTAATGTTTA | ATCTGACATA | ATGGCTAGGG | GAAATGAAGT | CTGCAGAATG |
| 20701 | GCTGTTTACG | GATGTTGTTG | TTGTTGTTGA | GATGAGGTCT | CACTATGTTG |
| 20751 | CCCAGGCTTG | AACTCCTGGC | CTCAAGCAGT | CCTCCTGCCT | TGACCTCCCA |
| 20801 | AAATGTTGAG | ATTACATGCA | TGAGCCATTG | CCAAAACGGC | TATTTGGATT |
| 20851 | GCTGTTAAGG | TTATTACATT | CTCTGTGTAG | TAAGACCTTG | AAGGAGAAGG |
| 20901 | ATTTGAGATC | AGGAGTTTAA | GAAAAAATGT | TAATCTAGGA | AGAGAGGATA |
| 20951 | ATTTCTGTGT | TTGGCCAGTT | GCAATGGCTC | ACGCCTGTAA | TCCCAGCGTT |
| 21001 | TTGGGAGGCC | GAGCTGGGCA | GATCACTTGA | GCTCAGGAGT | TTGAGAAGAG |
| 21051 | CCTGGGCAAC | ATGGTGAAGC | CCCGTCTCTA | CTGAAAATAC | AAAAATTGGC |
| 21101 | TGGGCGTGGT | GGCAGGTGCC | CACTGTAATT | TCAGCCACTC | AGGAGACTGA |
| 21151 | GGCAGGAGAA | TTGCTTGAAC | CTGGAAAGTG | GAGGATATGG | TGAGCCTAGA |
| 21201 | TCGCGCCACT | CTACTCCAGC | CTGGGAGACT | CCATCTCAAA | AAAAAAGAA  |
| 21251 | AACAAGATGC | TGAAATGAAG | TAATTACCAC | AGTCAATGTG | ATCCTATAAC |
| 21301 | TTTGTTTTCT | TTTAGAGATG | GGGTCTCCCT | CTGTCACCCA | GGCTGGAGTG |
| 21351 | CAGTGGTGCA | TCATAGCTTT | CTGCAGCCTC | CACCTCCTGG | GCTCAGGTGG |
| 21401 | TCCTCTTGCC | TCAGTGTTCC | GAGTAGTTAG | GACTGACTGC | AGGTGCATGC |
| 21451 | TGCTATGCCT | GGCTAACTTT | AAAATTTTTT | TGTAGAGGCG | GGGTCTTGCT |
| 21501 | ATGTTGCCTA | GGCTGGTCTC | CAACTCCTGA | TCTCAATCAG | TCCTCCTGCC |
| 21551 | TACCTTCCCA | GAGCGCTGGG | ATTACAGGTG | TGAGCCATCG | CACCTAGCCA |
| 21601 | ATCTCATAAC | ATTTTATGAC | TAGCAAACCT | AGTAGTTCTG | ATTCAGGCAT |
| 21651 | AAATCAGTTG | GTGGGGTTAT | ACAAGGTTGG | GTGAGTTTTT | CTAGATTTCT |
| 21701 | AAGAGACCAT | GTTGAAATAC | TTGGCTCTGT | CTCAGTAAGG | GATAGAGAGA |
| 21751 | AGCAAAGGTG | TGGGTAAAGG | TTATGAGCAG | ACATGTAAAG | GGGGCAAATT |
| 21801 | AAAGTGTTTA | GGGAAGGTGA | AACAGTTCCA | AATCATAATA | TAGATCCCAA |
| 21851 | GCCTCACCAA | GAAGTGAAGG | AGAGGGTAAT | TGTGCGATAG | TTCTCAGGAC |
| 21901 | TGAGACCTCA | AGGTGTAAGA | CGAATCTTTA | TTGTGGGTGG | TCCTCTTTGA |
| 21951 | GAAAAAGAAC | AAAAAAGAAA | AATGTGAAAT | GGGTGCTAAA | TTTCAGAGCA |
| 22001 | GAATATACAT | ATATGTGTAT | GTGTGTGTGT | ATATATATAT | ATATTTTTAT |

FIGURE 3, page 7 of 17



|       |            |            |            |            |            |
|-------|------------|------------|------------|------------|------------|
| 22051 | GTAGCAAATT | CCAGTTTACA | AAGGGCTTTT | ATGTGTTTTG | CATCATTACA |
| 22101 | ACAGTTCTGT | GATGATGTGG | ATGTGGTGGA | TATTGTAATT | CCACATCCCA |
| 22151 | GAGGATAAAA | TTGAGGCACA | CCATAAGGTA | GCTGACCAGA | GATCATGCAG |
| 22201 | TATATGGTAG | AATGGCAACT | TGAGGCCAGT | TCTTAGGTGT | GTTTGTGTGC |
| 22251 | CTTATTCAGA | ATGGAACAAT | GTGGTTTATT | GTACAAAAAT | TTAAAAATGA |
| 22301 | ATGTCGAAAA | GTAGAAATTA | TTACCCAAAT | CTCACCTCT  | GTGGTTGTTT |
| 22351 | GCTATGTGAT | CTTCCAGACT | CACATATACA | TGTAGATAAT | TTTTTTTTTT |
| 22401 | TTTTTTGAGA | CGGAGTCTTG | CTCTGTCGCC | CTGGCNNNNN | NNNNNNNNNN |
| 22451 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22501 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22551 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22601 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22651 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22701 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22751 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22801 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22851 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22901 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 22951 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23001 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23051 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23101 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23151 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23201 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23251 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23301 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23351 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23401 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23451 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23501 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23551 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23601 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23651 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23701 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23751 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23801 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23851 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23901 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 23951 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24001 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24051 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24101 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24151 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24201 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24251 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24301 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24351 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24401 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24451 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24501 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24551 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24601 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24651 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24701 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24751 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24801 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24851 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24901 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 24951 | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN |
| 25001 | NNNCAAAAAT | TAGCTGGGCG | TAGTGGTGGG | CGCCTGTGGT | CCCAGCTACT |
| 25051 | TGGGCAGCTG | AGGCAGGAGA | ATCGCTTGAA | CCTGGGAGGT | GGAGGTGCA  |
| 25101 | GTGAGCCAAG | ACCATGCCAC | TGCACTCCAG | CCTGGGCAAC | AGAGCAAGAC |
| 25151 | TCTTGTCTTA | AAAAACAAAA | AAGTGTACCC | AGTTGAGCTG | ATTCTTTATC |

FIGURE 3, page 8 of 17

|       |            |             |            |             |            |
|-------|------------|-------------|------------|-------------|------------|
| 25201 | TTTTTTTCAC | TGGAGAACTA  | AGTATACAGG | TGAGAAAAGA  | CGAGATATTT |
| 25251 | ATACCCGAGA | GAATTGATGG  | TGAAATCCAT | TTTTTTGGAT  | CAGAACTTCC |
| 25301 | CCAAACAGTG | TCCTTCAAAT  | AGGGTTCAGG | GGTGCTAAGA  | TATTTATCCC |
| 25351 | CTCAACCCTT | GGGGTTCACT  | CCAGTATGGC | ATATAAATAT  | TGTATCACTT |
| 25401 | TCTATGTGTG | GGGAGCAGTG  | CTCCAGGTGA | CCTTCCTTCC  | TTTCCTTCTA |
| 25451 | GGGGAGGAAG | TTGCCTTGTA  | CTGTGCCAAA | TATCTTCCTG  | ATATCATCAA |
| 25501 | AGATCAGAAG | GCCTACAAGG  | AAGGCAAGCT | ACAGAAGGTC  | TGTCTGCTTA |
| 25551 | CACCGCCCAT | TCCTCACTTG  | TGTAGGCTTT | TCCCTTGTTT  | TCTAGCCCTT |
| 25601 | GGGCTTTTCC | TTTCTTTTTG  | TCCTCTAGCT | GCTGCTGCTT  | ATTTACTCTT |
| 25651 | GAAGAATTCT | GTTCTTAAAA  | CGAGCTTATT | GGCCGCCTTT  | TAGACTTGCC |
| 25701 | TTATTATTCC | TAGGCCTCTG  | AGCTGTTTTT | ATCTGTGAGT  | GTCTCTTAGT |
| 25751 | GTGGTGGCTC | ACACTCTTAA  | TTTGTATTCC | ATCCTTGTGC  | TCAGGATTGT |
| 25801 | ATATAGGGAG | TTCATTTTGT  | ACTAGTCTTA | GACTATTTTG  | CTTATATTCA |
| 25851 | GGCTTTAGAA | GATGCCTTCT  | TGGCTATTGA | CGCCAAATTG  | ACCACTGAAG |
| 25901 | AAGTCATTAA | AGAGCTGGCA  | CAGATTGCAG | GGCGACCCAC  | TGAGGATGAA |
| 25951 | GATGAAAAAG | AAAAAGTAGC  | TGATGAAGAT | GATGGTGAGT  | GTGGCATCCC |
| 26001 | TTGTTTGAGG | GGAAATCAGC  | ATTTTAAGAA | ATATTCTTTA  | ATATTACTTA |
| 26051 | TCAATTCTAA | GATAGGATGG  | CTTTCTAGGG | ACCTGGGGAG  | TCCTTATGTT |
| 26101 | AAAGAAACCT | ATGATGTTCT  | CCTGCATTGT | ATGTGGTTAT  | GAAAAGGAGG |
| 26151 | GAGAGAATTA | TCTTCGTTGA  | GTGGCATCTG | AGCTGTAAGC  | ATTGTATATA |
| 26201 | CATTATCTTT | TGTCATTGTG  | ATGGGGTCTT | CCTGGTTCCT  | GCTAGTATTT |
| 26251 | ATGTGCTTTT | TTTTCCCCTC  | AAGACTGGAG | CAGTTATTAG  | CCCCAATAGC |
| 26301 | CAATCATTA  | GCCTAAATCC  | TAATTCACAG | TAGCATTGTG  | GGCTTCCTGG |
| 26351 | ATCCTCAGCC | AGAATAGGGT  | TTTTACAAC  | TAACAATAAA  | AAATGAGACG |
| 26401 | TCAGAGGGGA | AGTATAGTAA  | CTAGTGTTGT | TTTGATTAA   | AAGGGGATGA |
| 26451 | AACACAAAA  | CCAAAAGAAG  | TCTGTGGAGG | AGGAGGAGCT  | AGGGCATGTT |
| 26501 | CTTCTGAGAC | TTGAGCGAGA  | GGAACCTTGG | GAGTGGGAGG  | TTGTGGGGAA |
| 26551 | GTTAGAGGCT | GCAAGGGCTG  | TTGAGGTAGT | GAGAGGGACG  | GATCCCATGA |
| 26601 | GGAGTCTGGC | ATGGGGGCTC  | TGATTTAGCC | TCTTCCCTGC  | AGTGGACAAT |
| 26651 | GAGGAGGCTG | CACTGCTGCA  | TGAAGAGGCT | ACCATGACTA  | TTGAAGAGCT |
| 26701 | GCTGACACGC | TACGGGCAGA  | ACTGTCACAA | GGGCCCTCCC  | CACAGCAAAT |
| 26751 | CTGGAGGTGG | GACAGGCGAG  | GAACCAGGGT | CCCAGGGCCT  | CAATGGGGAG |
| 26801 | GCAGGACCTG | AGGACTCAAC  | TAGGGAAACT | CCTTCACAAG  | AAAATGGCCC |
| 26851 | CACAGCCAAG | GCCTACACAG  | GCTTTTCCTC | CAACTCGGAA  | CGTGGGACTG |
| 26901 | AGGCAGGCCA | AGTTGGTGAG  | CCTGGCATTC | CCACTGGTGA  | GGCTGGGCCT |
| 26951 | TCCTGCTCTT | CAGCCTCTGA  | CAAGCTGCCT | CGAGTTGCTA  | AGTCCAAGTT |
| 27001 | CTTTGAGGAC | AGTGAGGATG  | AGTCAGATGA | GGCGGAGGAA  | GAAGAGGAAG |
| 27051 | ACAGTGAGGT | AAGGGCCTGT  | GAGGGCAGGC | AGATGCTGAA  | GTTGCAGAGA |
| 27101 | GGTCCTGTTT | GGTTGCCGTC  | TGTAGTTTTT | AACTCTCTTT  | CCTTCTCCTA |
| 27151 | TTTTGACATC | ATCCCCCAAG  | ACCCACTGTA | TTCTAAGCTT  | TAGTCTTGAA |
| 27201 | TTCATTGAGC | TCCATCATCA  | CAGGTACCAT | TTGCCTTTTT  | ACCTCTTCCT |
| 27251 | TTGTTGGTAC | TATAACAAGC  | AGATCTAGTT | CTGGCTTTTC  | AGAGTCTGTC |
| 27301 | TCCTAGAGAG | AGAACAAGGA  | GATAGTTGTT | ACCTTGCGTA  | GTTGACTGTT |
| 27351 | TTCTTCTCTG | GAAAAATTTAT | TTTCTGGCCA | CAGTGCCTGA  | AAGATATTTT |
| 27401 | TGGCTGGCAG | CCCTTGCCCT  | GTCCTGGGCT | TTTTGCTAGT  | GAAGTCTAAG |
| 27451 | CCCAGTTCAG | GATGTCAGTT  | GTAATCATGC | TAGCCCTTTC  | CATCCCCCA  |
| 27501 | ATTTTCATGA | CCATATACTT  | GTATCTTTCA | GTGTTTTGAG  | GACCTGTGTT |
| 27551 | CAGTCAGGAC | CTCTTGATTC  | TGAGTATGAG | CTGTGGGGAG  | GGAGGGGATC |
| 27601 | ATCCCAGTCT | CAGCAGTCTG  | GGATCCTCCC | CCTGGCAGGA  | ATGCAGCGAG |
| 27651 | GAAGAGGATG | GCTACAGCAG  | TGAGGAGGCA | GAGAATGAGG  | AAGATGAGGA |
| 27701 | TGACACCGAG | GAGGCTGAAG  | AGGACGATGA | AGAAGAAGAA  | GAAGAGATGA |
| 27751 | TGGTGCCAGG | GATGGAAGGC  | AAAGAGGAGG | TGTGTGGGGA  | AGGGGAGCAA |
| 27801 | TGAGTCTTGA | AAAGCCACAA  | GGCAGGTGTG | AATCCCCTAA  | TTTTGATTTT |
| 27851 | GAGACAGGGG | ATCCCCCTGA  | TACTTTAGGA | TGGAAGTAAT  | AGTCATGGGG |
| 27901 | ATTTATTCTG | CAAGGGGAAT  | GAGATGGTAA | GCCTTTGGGG  | TTGAATTATC |
| 27951 | TAAAAACAAG | GGAGAGGGAG  | TGTGCTGCTG | TCTCTAGAAA  | GATGAAATGT |
| 28001 | GTGCTTCTCC | TGTTTGTTAA  | AGCTCTTTTG | GGGGTCCCAG  | TGAAAGCAAG |
| 28051 | CATAGGTGAA | CGATCAGGAG  | CACATCAGTG | AGGAACGCAT  | GTTCAGAAGC |
| 28101 | CCCCATGATG | CTCCTTTTCT  | TCCTCTTAAG | CCTGGCTCTG  | ACAGTGGTAC |
| 28151 | AACAGCGGTG | GTGGCCCTGA  | TACGAGGGAA | GCAGTTGATT  | GTAGCCAACG |
| 28201 | CAGGAGACTC | TCGCTGTGTG  | GTATCTGAGG | CTGGCAAAGC  | TTTAGACATG |
| 28251 | TCCTATGATC | ACAAACCAGA  | GGATGAAGTA | GAAC TAGCAC | GCATCAAGAA |
| 28301 | TGCTGGTGGC | AAGGTCACCA  | TGGATGGGCG | AGTCAACGGG  | GGCCTCAACC |

FIGURE 3, page 9 of 17

```

28351 TCTCCAGAGC CATTGGTAAG GGCCAAGAAA CTGGGAAAGA NNNNNNNNNN
28401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
28451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
28501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
28551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
28601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
28651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
28701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNATGGTGAA ACCCCATCTC
28751 TACTAAAAAA AAAAAAAAT ACAAATAA GCTGGGCATG GTAGTGCAAG
28801 CCTGTAATCC CAGCTACTCA GGAGGCTGAG GCAGGAGAAC CGCTTGAATC
28851 CGGGAGGCGG AGGTTGTAGT GAGCCGAGAT CGTGCCATTA CACTCCAGCC
28901 TGGGCTACAA GAGTGAAACT CCGTCTCAAA AAAAAAAAAA CAAAAAGAC
28951 TTAAATAAAA AAGACCAGTG AGTGACTTTC TTAAGGTTCA GCAGTCTGGT
29001 GGCAGGGTTG AAAC TAGAAA AACTAGGACT TAGGACTCAG TTCCCCATTC
29051 CACTAGATTA TGGAAC TTTG TAAAGAAGGG AAATGAATGG CAAGGTTTGA
29101 CCTGCCACAA ACACAAGTCT GTGGGAAGTA TCCAAACTGC TCATCAACCA
29151 TTCCTTTACT CCAGGGGACC ACTTCTATAA GAGAAACAAG AACCTGCCAC
29201 CTGAGGAACA GATGATTTCA GCCCTTCCTG ACATCAAGGT GCTGACTCTC
29251 ACTGACGACC ATGAATTCAT GGTCATTGCC TGTGATGGCA TCTGGTGAGC
29301 ACTGGCAGAA TGCCCTAAAT TCCCCTTTCT GCAGCATGTC TTCTCTTATA
29351 GGA CTCAGGG CACCTCTAGG ATTAGAGCCT AGGCAGACCT AGGCCTCTTG
29401 GTGGGTGAAG AGCACCCAGA CTAAGGCAGA GCTGAGAATT TCTGTAGTTA
29451 TTTACTACTGG CCTGGGCCAC CACCTCTGTC CATACTCCTC TACGCTGCCT
29501 TAGTGAGACT GGAAGATTCT GACTGTTGTT CTTGACCCCA GGAATGTGAT
29551 GAGCAGCCAG GAAGTTGTAG ATTTCAATCA ATCAAAGATC AGCCAGCGTG
29601 ATGAAAATGG GGAGCTTCGG TTATTGTCAT CCATTGTGGA AGAGGTGAGT
29651 ACCAGGGTGG AGAAGAGAGG GTGTCTGGTC TGCACAGCCA GGGTT (SEQ ID NO:3)

```

**FEATURES:**

```

Start:      2134
Exon:       2134-2326
Intron:     2327-25451
Exon:       25452-25537
Intron:     25538-25851
Exon:       25852-25984
Intron:     25985-29164
Exon:       29165-29294
Intron:     29295-29541
Exon:       29542-29647
Stop:

```

**CHROMOSOME MAP POSITION:**

Chromosome 2

**ALLELIC VARIANTS (SNPs):**

| DNA      |       |       |
|----------|-------|-------|
| Position | Major | Minor |
| 2944     | A     | -     |
| 3232     | G     | A     |
| 7236     | T     | A     |
| 8187     | T     | A G   |
| 8187     | T     | - G   |
| 8187     | -     | A G   |
| 9172     | C     | A     |
| 9173     | T     | -     |
| 9180     | T     | -     |
| 13351    | T     | C     |
| 15100    | G     | C     |
| 15255    | A     | C     |
| 18996    | A     | C     |
| 18996    | A     | C     |
| 18996    | A     | C     |



|       |   |   |
|-------|---|---|
| 20532 | C | T |
| 21118 | G | T |
| 21211 | C | G |
| 21212 | T | C |
| 21238 | - | A |
| 26483 | G | A |

Context:

DNA  
Position  
2944

AGATGTGAATAACTTTTTAAATTTTAATATTTTAAATATCTGATGTGGGAAGCCTCTTTT  
GGCTAGGAGTTTGACAGTGAAAGGAACCCCGGGCAGAGTCTGTTTCACATTTTGGTTGCC  
TGGCCTTGGGCTCTTGCATGTTAATTTTCAAGAGGCTGGACCCGACCTCCAGGAGTTGTCAC  
TCATTTGCACTCTTTTCAAGGCTTTTACTAACTTCGGAAAACCTGAATTATGTCAGTCC  
CTAGGTTTTCTTTTATTATGCTTCGTTTTCTTTCTCTCTTAGGATTTCTCTAAACTT  
[A, -]  
ATCAGTAATTCTCATTTGCCTGTAATTGTAGTTTATAGTTTCTTTGATTGTAAGTCATTT  
CATCAATTTTTTTTATCACAACCTACCTACAAAGGGCTTTTCTAGAAAATTTTACTCTGG  
ACAAAAGGGGAAAAGAAAATATTGGGGGGAAAGTAGTAGTATTAGGTAAACTTGATGT  
GAAACTACAAAAGAGAAGAGGGAAAACGCGGTAGGGAGGAAAGGGAGGAAGACGGGTTA  
ACCGTGGCTTTGTGAAGAGCATTCTGAAGTCTAGGCAAAGGGCCAGGGAAATACTCTGT

3232 TCTCTAAACTTAATCAGTAATTCTCATTTGCCTGTAATTGTAGTTTATAGTTTCTTTGA  
TTGTAAGTCATTTTATCAATTTTTTTTATCACAACCTACCTACAAAGGGCTTTTCTAGAA  
AATTTTACTCTGGACAAAAGGGGAAAAGAAAATATTGGGGGGAAAGTAGTAGTATTAGG  
TAAACTTGATGTGAACTACAAAAGAGAAGAGGGAAAACGCGGTAGGGAGGAAAGGGA  
GGAAGACGGGTAAACCGTGGCTTTGTGAAGAGCATTCTGAAGTCTAGGCAAAGGGCCAG  
[G, A]  
GAAATACTCTGTCTGGTATTGAGGGTTTCTCCACCTACCGGGTGGGCTTCAGGTAAACAGC  
GAAATACTGTCTCCCTTGGGAATTGTTTTCAGATCCCTCGCTCCTCCTGTGGTTAGCTCTG  
GAATGCCAGTATGAACCTCAATGTTTTGTTTTCCGATTCAAATTTTATATTCACTAAGTGA  
CCTTAATAACAATTTTACAATTAGGTATAAAATTTTCAAGATCCTAGTGTATCCTATAGTT  
CATCTCATCTGCTTTGGCTCCCTTTTTTTTTTTTTTGTAGACAGAATTTTCGCTCTTGTTGCC

7236 GACAGGATTTCACTATGTTGGCCAGGCTGGTCTCAAGACTCCTGACGTTGTGATCCACCT  
GCCTCGGCCTCCCAAAGTGCTGGGATTACAGATGTGAGCTACCGTGCCCGGCCAGGAATT  
TTTTGTGCTATAAATCATATTTTCTTTTATTAAAGGCAGTGTCAATATCTATAGTATAA  
TTTTGAGGAGGCTGGCTATTTATTGCTGTGTAGAAGCTGGCTTATTAGTGGTCAAGGGGT  
CATCTAGAATTGACTATAAAGATAGTATTGAGCAGAAAATTCTTAAATACCTGCATATT  
[T, A]  
GTTTCAGTCATTAAATTAATGGAAAAATATAAAAAGAAATATCACAAGTATGCTATGGG  
GTTCTACCTTAGGGCATTGAAGGTTGAAAAACATTTATTTTCTTATCTTCAGATTAGCAT  
CTCATATCAATAGCCAATAGCTTAAAGCGCTTTTACTTACTAAACCAGGTCAGAATTTT  
TCTCTCTCTCTTTTTTTTTTTTTTAGACAGAGTCTCTGTCACCCAGTCTAGAGTGCAGTG  
GTGTGATCTCGGCTCACTGGCCTTTCTGTTTAAAGTGAATCTCCGGCCTCAGCCTCCTGAG

8187 GTTGCCCAGGCTGGAGTGCAATGGTGCCATCTCGGCTCACTGCAGCTTCTGCCTCCTGGG  
TTCTAGCAATTTGCCTACCCAGCCTCCCAAGTAGCTGGGATTACAGGCACTCGCCACCA  
TGCCCAGCTAACTTTTTTTGTATTTTATAGTAGAGAAGGGGTTTTCCCATGCTGGCGAGTCT  
GGTCTTGAACCTTCTGGCCTCAAGTGATCTGCCCAGCTCAGCCTCCCAAAGTGCTGGAATT  
ACAGGCATGAGCCACTGTGCCCAGCCTTTTTATTTTTTTAAATTATTTATGTATTTATTTT  
[T, A, G]  
AGACAGGATCTCACTCTTGCCCATGCTTGAGTGGTAAGGAGTATGGGATTTCTTGTCGG  
GTTCTTTCACATCCTCACTACACTTATCTGCCTTTACAGTGGCTCGATCATAGTTCACTG  
CATAGCCTTCTGGGCTCAAGGGGTCTTCCAGCCTCAGCCTAATATAGGCACATGCCACCA  
TGACTGGCTAATTTTTTTTTTTTTTAAGTTTTTTTTTGTAGAGATAGGGCCTTGCAAGTGTG  
CCCAGGCTGAGGAATTTTATTTATGTTTATTTTATTTATTTATTTATTTATTTATTTATTT

8187 GTTGCCCAGGCTGGAGTGCAATGGTGCCATCTCGGCTCACTGCAGCTTCTGCCTCCTGGG  
TTCTAGCAATTTGCCTACCCAGCCTCCCAAGTAGCTGGGATTACAGGCACTCGCCACCA  
TGCCCAGCTAACTTTTTTTGTATTTTATAGTAGAGAAGGGGTTTTCCCATGCTGGCGAGTCT  
GGTCTTGAACCTTCTGGCCTCAAGTGATCTGCCCAGCTCAGCCTCCCAAAGTGCTGGAATT

FIGURE 3, page 11 of 17

ACAGGCATGAGCCACTGTGCCCCGGCCTTTTTATTTTTTAAATTATTTATGTATTTATTTT  
[T, -, G]  
AGACAGGATCTCACTCTTGCCCATGCTTGAGTGGTAAGGAGTATGGGATTTCTTGTGCCG  
GTTCCCTTCACATCCTCACTACACTTATCTGCCTTTACAGTGGCTCGATCATAGTTCACTG  
CATAGCCTTCTGGGCTCAAGGGGTCTTCCAGCCTCAGCCTAATATAGGCACATGCCACCA  
TGACTGGCTAATTTTTTTTTTTTAAAGTTTTTTTTTTGTAGAGATAGGGCCTTGCAGTGTTG  
CCCAGGCTGAGGAATTTTATTTATGTTTATTTTATTTATTTATTTATTTATTTATTTATT

8187 GTTGCCCAGGCTGGAGTGCAATGGTGCCATCTCGGCTCACTGCAGCTTCTGCCTCCTGGG  
TTCTAGCAATTTGCCTACCCAGCCTCCCAAGTAGCTGGGATTACAGGCACTCGCCACCA  
TGCCCAGCTAACTTTTTTTGTATTTTTTAGTAGAGAAGGGGTTTTCCCATGCTGGCGAGTCT  
GGTCTTGAACCTTCTGGCCTCAAGTGATCTGCCCCGCTCAGCCTCCCAAAGTGCTGGAATT  
ACAGGCATGAGCCACTGTGCCCCGGCCTTTTTATTTTTTAAATTATTTATGTATTTATTTT  
[-, A, G]  
AGACAGGATCTCACTCTTGCCCATGCTTGAGTGGTAAGGAGTATGGGATTTCTTGTGCCG  
GTTCCCTTCACATCCTCACTACACTTATCTGCCTTTACAGTGGCTCGATCATAGTTCACTG  
CATAGCCTTCTGGGCTCAAGGGGTCTTCCAGCCTCAGCCTAATATAGGCACATGCCACCA  
TGACTGGCTAATTTTTTTTTTTTAAAGTTTTTTTTTTGTAGAGATAGGGCCTTGCAGTGTTG  
CCCAGGCTGAGGAATTTTATTTATGTTTATTTTATTTATTTATTTATTTATTTATTTATT

9172 AAGGATATTTTACACAAAGTTCTATTGTCTGATTCCCTAGCAGTTGTTACTACTGTTTCC  
CTAACCTCTAATCTTCTATTGGGTTATTAGTCTTAGAATTGAATTTTGAGAGGTAAGGGC  
TTGAATTTGAACATAGAAATTTATACAGGTCTGATCAGTAGTTCTTGACATTGTATTATC  
TGGAACAAATCTTTAGAACTGAGCTTAAGATGTTTAAATGACATTTTGTAGACAGAGTAT  
GATTTCAAGTGTAGTTGTTTTTGTCTTTCTAGATCTAGTTCAGAGATGAAGTATATCAA  
[C, A]  
TTTTTTTTTCCTTTTTGACCCAATGCTAGCAGAAAAACAACACCTTTTAATCATATTTAGT  
ATTTGAAAATGTGTATACAGGTTCCTTTTATTTTATTTATTTCTTACAGGTTCCTTTTT  
AATCAGCTTTATTGAGATAGAGTTCATATACTGTATGGTTCATACCACATATGGTTCATA  
TACCATACAGTATATGAACTCACTTTAAGAGTATAATTCAGTGGGTTTAAGGGTATAATT  
CATTCATTTTAAGGGTATAATTCAGTGGCTTTTAGTATATTTTCTTTTTTTTCTTCTTTT

9173 AGGATATTTTACACAAAGTTCTATTGTCTGATTCCCTAGCAGTTGTTACTACTGTTTCCC  
TAACCTCTAATCTTCTATTGGGTTATTAGTCTTAGAATTGAATTTTGAGAGGTAAGGGCT  
TGAATTTGAACATAGAAATTTATACAGGTCTGATCAGTAGTTCTTGACATTGTATTATCT  
GGAAACAAATCTTTAGAACTGAGCTTAAGATGTTTAAATGACATTTTGTAGACAGAGTATG  
ATTTCAAGTGTAGTTGTTTTTGTCTTTCTAGATCTAGTTCAGAGATGAAGTATATCAAC  
[T, -]  
TTTTTTTTTCCTTTTTGACCCAATGCTAGCAGAAAAACAACACCTTTTAATCATATTTAGTA  
TTTGAAAATGTGTATACAGGTTCCTTTTATTTTATTTATTTCTTACAGGTTCCTTTTTA  
ATCAGCTTTATTGAGATAGAGTTCATATACTGTATGGTTCATACCACATATGGTTCATAT  
ACCATACAGTATATGAACTCACTTTAAGAGTATAATTCAGTGGGTTTAAGGGTATAATTC  
ATTCATTTTAAGGGTATAATTCAGTGGCTTTTAGTATATTTTCTTTTTTTTCTTCTTTT

9180 TTTACACAAAGTTCTATTGTCTGATTCCCTAGCAGTTGTTACTACTGTTTCCCTAACCTC  
TAATCTTCTATTGGGTTATTAGTCTTAGAATTGAATTTTGAGAGGTAAGGGCTTGAATTT  
GAACATAGAAATTTATACAGGTCTGATCAGTAGTTCTTGACATTGTATTATCTGGAAACA  
AATCTTTAGAACTGAGCTTAAGATGTTTAAATGACATTTTGTAGACAGAGTATGATTTTCA  
GTAGTTGTTTTTGTCTTTCTAGATCTAGTTCAGAGATGAAGTATATCAACTTTTTTT  
[T, -]  
CCTTTTTTGACCCAATGCTAGCAGAAAAACAACACCTTTTAATCATATTTAGTATTTGAAA  
ATGTGTATACAGGTTCCTTTTATTTTATTTATTTCTTACAGGTTCCTTTTTAATCAGCT  
TTATTGAGATAGAGTTCATATACTGTATGGTTCATACCACATATGGTTCATATACCATAC  
AGTATATGAACTCACTTTAAGAGTATAATTCAGTGGGTTTAAGGGTATAATTCATTCATT  
TTAAGGGTATAATTCAGTGGCTTTTAGTATATTTTCTTTTTTTTCTTCTTTTTTCTTTT

13351 TTTTCCCCCATTTGAATTATCTTGGCATCATTGTCAGAGATAAATTGACCGTAAATGTGAG  
GGTTTTATTTCTGAACTCTCAAGTCCATTTTATTGGTCTACATGTCCCTATGCCAGTAAT  
ACACTATCTTGGTTACTGTAGCTTTTTTAGTACGTTTTGAAATGTTTTTAAATTTGTTTT  
TCATCTAAATTTTAGGATTAATTTGTCAATTTCTGCACAAAGGCACCTGGGTTTCTATA  
GGGGTTATGCAGAATCTGTAGATCAACTGGGGGAGTATTACAGGCATGAGCCACCGTGCC  
[T, C]  
GGCTGACTGAGTTTTTCATAGATGTACTCTATCAGGTTTAGGAAGTTCCCTTTTATTCCT

FIGURE 3, page 12 of 17

AGGTTGTTGAGTCTATTTTATATTACTTTTTTAGAGACAGTCTTGCTCTGTCCCTCAGGC  
TGTAGCACAGTGGCTCAATCATAGCTCACTGCAGCCTTGAACCTAGGTTCAAGAGATC  
CGCCTGCCTCAGCCTTCTTAGTAGCTGGGATTACATGCATGCACCACCATACTGGGCTAA  
TTTTTTAAAATTTTTTATAGAGACAGGGTCTTATTACTATGTTGCCCAGACTGGCATTGA

15100 GTTGTGTACCCAGTTTCCAGAACTCTTCATCCTACAGAACTGAAACTCCATACCCATTAA  
ATGAGTCCCCATTCTCTTTCCCCCAGCTCATGGCAAACAGCATTCTATTTTCAGTCTCTA  
TGAATTTGATTAGTTTAGATACTTCATACTGTAAGTGGAAATCATATGGTATTTGTCTTTT  
AGTGACTGCCTAATTTAAAAAAATTTTTTTGAGACGGAGTCCTGCTCTGTGCCCCAGGC  
TGGAGTGCAGTGGCACCATCTCTGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATT  
[G, C]  
TCCTGCCTCGGCCTCCCACGTAGCTGGGATTACAGGTGCTCGCCACAACACCCGGCTAAT  
TTTTGTATTTTATAGGTAGAGACTGGGTTTCAACATGTTGGCCAGGCTGGTCTCGAACTCC  
TGACCTCAAATTATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCC  
ACTGTGCCCAGCCTCCATGTTGTTTTTCAACAACCTGTATCATTTACATTTCCACCAAC  
AGTACACAAGAATTTTCAAGTTTCTCCACATCCTTGCTAGCAGTTGTTATTATCTGTTTTTT

15255 TGGAATCATATGGTATTTGTCTTTTAGTGACTGCCTAATTTAAAAAAATTTTTTTGAGA  
CGGAGTCCTGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCACCATCTCTGCTCACTGCAAC  
CTCCACCTCCCAGGTTCAAGTGATTGTCCTGCTCGGCCTCCCACGTAGCTGGGATTACA  
GGTGCTCGCCACAACACCCGGCTAATTTTTGTATTTTATAGGTAGAGACTGGGTTTCACCA  
TGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAAATTATCCACCTGCCTTGGCCTCCCA  
[A, C]  
AGTGCTGGGATTACAGGCGTGAGCCACTGTGCCCAGCCTCCATGTTGTTTTTCAACAAC  
CTGTATCATTTACATTTCCACCAACAGTACACAAGAATTTCAAGTTTCTCCACATCCTTGC  
TAGCAGTTGTTATTATCTGTTTTTTTTTAATGGTTTCTTTTTTCTTTTTTTTTTTTT  
TTTTTTGAGACGGTCTTATTCTTGCTCATGCTGGAGTGCAGTGGTGCAATGTGATAGCTC  
ACTGCAGCCTCAACCTCTGGGCTCAAGCAGTCCTGCCACCTCAGCCTCCACATAGGTGGG

18996 GTAGTGGCTCATGCCTAAATCTCAGCACTTTGGGAGGCTGAGGCGGGAGGATCACCTGAG  
GTCGGGAGTTTGAGACCAGCCTGGGCAACATGGTGAAACCCCGTGTCTACTAAAAATATA  
AAAAGTAGCCAGTCATGGTGGCACACACCTGTAATCCCGGCTACTTGGGAGGCTGAGGCA  
TGAGAATCACTTGAACCTGGGCTGCGGAGGTTGTAGTAAGCTGAGATCATGCCACTGTAC  
TCCAGCCTGGGTGACAGGGAGACATTCTCTCTCTCAAAAAAAAAAAAAAAAAACAAAAAAA  
[A, C]  
AAAACAAACCAACAAAAACAAAGTAATCCAGGAACAACAACATGATGAAGGACTGCATGCA  
GGACTCAGTGATGGATGGTGGAGACAGCCAGGAAGTTAAGCATGACTCTGGTATTAAGT  
GTTGTCTGGGAGAGTTAAGATTCCATTTACAGAAATAAGACCTGTAGGGGAAGCTCTTGA  
TTTTTTTTTTTTTGCAGACTGCTGATTTCTGATTACATGTGTTAAGTTTGAGGTATAGA  
GAGAAAGAACATCCTGGCCGGGTGCAGTGGCTCACACCCGTAATCCCAGCACTTTGGGAG

18996 GTAGTGGCTCATGCCTAAATCTCAGCACTTTGGGAGGCTGAGGCGGGAGGATCACCTGAG  
GTCGGGAGTTTGAGACCAGCCTGGGCAACATGGTGAAACCCCGTGTCTACTAAAAATATA  
AAAAGTAGCCAGTCATGGTGGCACACACCTGTAATCCCGGCTACTTGGGAGGCTGAGGCA  
TGAGAATCACTTGAACCTGGGCTGCGGAGGTTGTAGTAAGCTGAGATCATGCCACTGTAC  
TCCAGCCTGGGTGACAGGGAGACATTCTCTCTCTCAAAAAAAAAAAAAAAAAACAAAAAAA  
[A, C]  
AAAACAAACCAACAAAAACAAAGTAATCCAGGAACAACAACATGATGAAGGACTGCATGCA  
GGACTCAGTGATGGATGGTGGAGACAGCCAGGAAGTTAAGCATGACTCTGGTATTAAGT  
GTTGTCTGGGAGAGTTAAGATTCCATTTACAGAAATAAGACCTGTAGGGGAAGCTCTTGA  
TTTTTTTTTTTTTGCAGACTGCTGATTTCTGATTACATGTGTTAAGTTTGAGGTATAGA  
GAGAAAGAACATCCTGGCCGGGTGCAGTGGCTCACACCCGTAATCCCAGCACTTTGGGAG

18996 GTAGTGGCTCATGCCTAAATCTCAGCACTTTGGGAGGCTGAGGCGGGAGGATCACCTGAG  
GTCGGGAGTTTGAGACCAGCCTGGGCAACATGGTGAAACCCCGTGTCTACTAAAAATATA  
AAAAGTAGCCAGTCATGGTGGCACACACCTGTAATCCCGGCTACTTGGGAGGCTGAGGCA  
TGAGAATCACTTGAACCTGGGCTGCGGAGGTTGTAGTAAGCTGAGATCATGCCACTGTAC  
TCCAGCCTGGGTGACAGGGAGACATTCTCTCTCTCAAAAAAAAAAAAAAAAAACAAAAAAA  
[A, C]  
AAAACAAACCAACAAAAACAAAGTAATCCAGGAACAACAACATGATGAAGGACTGCATGCA  
GGACTCAGTGATGGATGGTGGAGACAGCCAGGAAGTTAAGCATGACTCTGGTATTAAGT  
GTTGTCTGGGAGAGTTAAGATTCCATTTACAGAAATAAGACCTGTAGGGGAAGCTCTTGA  
TTTTTTTTTTTTTGCAGACTGCTGATTTCTGATTACATGTGTTAAGTTTGAGGTATAGA

FIGURE 3, page 13 of 17

20532  
21118  
21211  
21212  
21238  
26483

GAGAAAGAACATCCTGGCCGGGTGCAGTGGCTCACACCCGTAATCCCAGCACTTTGGGAG  
20532 TAGCCATCATGGAATACTATGCACCCATTAAAAATAATGAAATAATGTCTTTGCAACAAC  
ATGAATGTAGCTGGAGGGCATTATCCTAAGCAAATAACACAGAAACAGAAAACCAAATA  
CTGCGTGTTCTCACGCAGTGAGAGTGGGAGCTGAACATCAAGTACACATGGATGTAAAGA  
TGGCAACAATAGACATGGGTCTACTAGAGGTGGTGGTGCGGCAGGGTGGGGGTGGGGGT  
GTGTGGCAGAGGAACAGCTGAAAACTACCTATTTGATACTATACCCAGCACCTGGGAAA  
[C,T]  
GGGTTCAGTCATACCCCAAACCTCAGCATCACACAGTATACCTTTCTAACAACTTACAC  
ATGTATTCTGTGATTCTAAAATAAACATTGAAAAATAAAAAAAACTGACATGGTTTGT  
ACTGTTTAATCTGACATAATGGCTAGGGGAAATGAAGTCTGCAGAATGGCTGTTTACGGA  
TGTTGTTGTTGTTGTTGAGATGAGGTCTCACTATGTTGCCCAGGCTTGAACCTGGCCT  
CAAGCAGTCCTCCTGCCTTGACCTCCCAAATGTTGAGATTACATGCATGAGCCATTGCC  
21118 GCATGAGCCATTGCCAAAACGGCTATTTGGATTGCTGTAAAGGTTATTACATTCTCTGTG  
TAGTAAGACCTTGAAGGAGAAGGATTTGAGATCAGGAGTTTAAGAAAAAATGTTAATCTA  
GGAAGAGAGGATAATTTCTGTGTTTGGCCAGTTGCAATGGCTCACGCCTGTAATCCCAGC  
GTTTTGGGAGGCCGAGCTGGGCAGATCACTTGAGCTCAGGAGTTTGAGAAGAGCCTGGGC  
AACATGGTGAAGCCCCGTCTCTACTGAAAATACAAAAATTGGCTGGGCGTGGTGGCAGGT  
[G,T]  
CCCCTGTAATTTGAGCCACTCAGGAGACTGAGGCAGGAGAATTGCTTGAACCTGGAAAG  
TGGAGGATATGGTGAGCCTAGATCGCGCCACTCTACTCCAGCCTGGGAGACTCCATCTCA  
AAAAAAAAGAAAACAAGATGCTGAAATGAAGTAATTACCACAGTCAATGTGATCCTATA  
ACTTTGTTTTCTTTTAGAGATGGGGTCTCCCTCTGTCAACCAGGCTGGAGTGCAGTGGTG  
CATCATAGCTTTCTGCAGCCTCCACCTCCTGGGCTCAGGTGGTCCTCTTGCCTCAGTGTT  
21211 AGGAGTTTAAGAAAAAATGTTAATCTAGGAAGAGAGGATAATTTCTGTGTTTGGCCAGTT  
GCAATGGCTCACGCCTGTAATCCCAGCGTTTTGGGAGGCCGAGCTGGGCAGATCACTTGA  
GCTCAGGAGTTTGAGAAGAGCCTGGGCAACATGGTGAAGCCCCGTCTCTACTGAAAATAC  
AAAAATTGGCTGGGCGTGGTGGCAGGTGCCCACTGTAATTTGAGCCACTCAGGAGACTGA  
GGCAGGAGAATTGCTTGAACCTGGAAAGTGGAGGATATGGTGAGCCTAGATCGCGCCACT  
[C,G]  
TACTCCAGCCTGGGAGACTCCATCTCAAAAAAAAAGAAAACAAGATGCTGAAATGAAGT  
AATTACCACAGTCAATGTGATCCTATAACTTTGTTTTCTTTTAGAGATGGGGTCTCCCTC  
TGTCACCCAGGCTGGAGTGCAGTGGTGCATCATAGCTTTCTGCAGCCTCCACCTCCTGGG  
CTCAGGTGGTCCTCTTGCCTCAGTGTTCCGAGTAGTTAGGACTGACTGCAGGTGCATGCT  
GCTATGCCTGGCTAACTTTAAAATTTTTTTGTAGAGGCGGGGTCTTGCTATGTTGCCTAG  
21212 GGAGTTTAAGAAAAAATGTTAATCTAGGAAGAGAGGATAATTTCTGTGTTTGGCCAGTTG  
CAATGGCTCACGCCTGTAATCCCAGCGTTTTGGGAGGCCGAGCTGGGCAGATCACTTGAG  
CTCAGGAGTTTGAGAAGAGCCTGGGCAACATGGTGAAGCCCCGTCTCTACTGAAAATACA  
AAAATTGGCTGGGCGTGGTGGCAGGTGCCCACTGTAATTTGAGCCACTCAGGAGACTGAG  
GCAGGAGAATTGCTTGAACCTGGAAAGTGGAGGATATGGTGAGCCTAGATCGCGCCACTC  
[T,C]  
ACTCCAGCCTGGGAGACTCCATCTCAAAAAAAAAGAAAACAAGATGCTGAAATGAAGTA  
ATTACCACAGTCAATGTGATCCTATAACTTTGTTTTCTTTTAGAGATGGGGTCTCCCTCT  
GTCACCCAGGCTGGAGTGCAGTGGTGCATCATAGCTTTCTGCAGCCTCCACCTCCTGGGC  
TCAGGTGGTCCTCTTGCCTCAGTGTTCCGAGTAGTTAGGACTGACTGCAGGTGCATGCTG  
CTATGCCTGGCTAACTTTAAAATTTTTTTGTAGAGGCGGGGTCTTGCTATGTTGCCTAGG  
21238 GGAAGAGAGGATAATTTCTGTGTTTGGCCAGTTGCAATGGCTCACGCCTGTAATCCCAGC  
GTTTTGGGAGGCCGAGCTGGGCAGATCACTTGAGCTCAGGAGTTTGAGAAGAGCCTGGGC  
AACATGGTGAAGCCCCGTCTCTACTGAAAATACAAAAATTGGCTGGGCGTGGTGGCAGGT  
GCCCACTGTAATTTGAGCCACTCAGGAGACTGAGGCAGGAGAATTGCTTGAACCTGGAAA  
GTGGAGGATATGGTGAGCCTAGATCGCGCCACTCTACTCCAGCCTGGGAGACTCCATCTC  
[-,A]  
AAAAAAAAGAAAACAAGATGCTGAAATGAAGTAATTACCACAGTCAATGTGATCCTATA  
ACTTTGTTTTCTTTTAGAGATGGGGTCTCCCTCTGTCAACCAGGCTGGAGTGCAGTGGTG  
CATCATAGCTTTCTGCAGCCTCCACCTCCTGGGCTCAGGTGGTCCTCTTGCCTCAGTGTT  
CCGAGTAGTTAGGACTGACTGCAGGTGCATGCTGCTATGCCTGGCTAACTTTAAAATTTT  
TTTGTAGAGGCGGGGTCTTGCTATGTTGCCTAGGCTGGTCTCCAACCTCCTGATCTCAATC  
26483 CTGTAAGCATTGTATATACATTATCTTTTGTCATTGTGATGGGGTCTTCCTGGTTCCTGC

FIGURE 3, page 14 of 17

TAGTATTTATGTGCTTTTTTTTCCCCTCAAGACTGGAGCAGTTATTAGCCCCAATAGCCA  
ATCATTAAGCCTAAATCCTAATTCACAGTAGCATTGTGGGCTTCCTGGATCCTCAGCCAG  
AATAGGGTTTTTACAACCTAACAATAAAAAATGAGACGTCAGAGGGGAAGTATAGTAACT  
AGTGTTGTTTTGATTAAGAAGGGGATGAAACACAAAAACCAAAGAAGTCTGTGGAGGAG  
[G,A]  
AGGAGCTAGGGCATGTTCTTCTGAGACTTGAGCGAGAGGAACCTTGGGAGTGGGAGGTTG  
TGGGGAAGTTAGAGGCTGCAAGGGCTGTTGAGGTAGTGAGAGGGACGGATCCCATGAGGA  
GTCTGGCATGGGGGCTCTGATTTAGCCTCTTCCCTGCAGTGGACAATGAGGAGGCTGCAC  
TGCTGCATGAAGAGGCTACCATGACTATTGAAGAGCTGCTGACACGCTACGGGCAGAACT  
GTCACAAGGGCCCTCCCCACAGCAAATCTGGAGGTGGGACAGGCGAGGAACCAGGGTCCC

GENEWISE ALIGNMENT:

|                |       |   |                         |
|----------------|-------|---|-------------------------|
| gi 4505999 ref | 1     | MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPEL   |                         |
|                |       | MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSME            |                         |
| genomic        | 2134  | aggtctccaagattggggggccccctgttgacgtcgtaggaacgcagt    |                         |
|                |       | tgcatacactagcgagtgccgtctcagtcctaggggtctatggagctgg   |                         |
|                |       | gtccccgccggggccgcgcccgcgggccccccgacgccccgggggggcacg |                         |
| gi 4505999 ref | 50    | DSETAMFSVYDGHGG                                     | EEVALYCA                |
|                |       | ++ + G  | EEVALYCA                |
| genomic        | 2281  | PLRGGNLTEKESGGW - :G[ggg]                           | EEVALYCA                |
|                |       | cccgacagagaggtGGTCTCTC Intron 1 TAGGGgggggtttg      |                         |
|                |       | ctgggatcaaagggg <1-----[2327 :25451]-1> aatctagc    |                         |
|                |       | gggcgtgggagcgag gatcgctc                            |                         |
| gi 4505999 ref | 73    | KYLPDIKDQKAYKEGKLQK                                 | ALEDAF                  |
|                |       | KYLPDIKDQKAYKEGKLQK                                 | ALEDAF                  |
| genomic        | 25478 | KYLPDIKDQKAYKEGKLQK                                 | ALEDAF                  |
|                |       | atccgaaagcagtaggaccaGTCTGTC Intron 2 CAGgtgggt      |                         |
|                |       | aatcattaaacaaagataa<0-----[25538:25851]-0>ctaact    |                         |
|                |       | attttccatggccgacgagg taatcc                         |                         |
| gi 4505999 ref | 99    | LAIKAKLTTEEVIKELAQIAGRPTEDDEDEKEKVADEDD             |                         |
|                |       | LAIKAKLTTEEVIKELAQIAGRPTEDDEDEKEKVADEDD             |                         |
| genomic        | 25870 | LAIKAKLTTEEVIKELAQIAGRPTEDDEDEKEKVADEDD             |                         |
|                |       | tgaggataaggggaagcgcaggccagggggagagggggg             |                         |
|                |       | tctacatccaattaatcatcgcccaaaaaaatcaaaa               |                         |
|                |       | gttccagctaactaggagtagactgtataaaaattatt              |                         |
| gi 4505999 ref | 137   |   | DNEEAALLHEEATMTIEELLTRY |
|                |       |   | DNEEAALLHEEATMTIEELLTRY |
| genomic        | 25984 | V:V[gtg]  | DNEEAALLHEEATMTIEELLTRY |
|                |       | GGTGAGTG Intron 3 CAGTGgagggggccccgggaaaaggccact    |                         |
|                |       | <1-----[25985:26642]-1> aaaaccttaaactctaattcga      |                         |
|                |       | ctggtaggtagtcgttagggacc                             |                         |
| gi 4505999 ref | 161   | GQNCHKGPPHSKSGGGTGEEPGSQGLNGEAGPEDSTRETSPQENGPTAK   |                         |
|                |       | GQNCHKGPPHSKSGGGTGEEPGSQGLNGEAGPEDSTRETSPQENGPTAK   |                         |
| genomic        | 26714 | GQNCHKGPPHSKSGGGTGEEPGSQGLNGEAGPEDSTRETSPQENGPTAK   |                         |
|                |       | gcatcagcccaatgggagggcgctcgaggggcggttaagactcgagcaga  |                         |
|                |       | gaagaagccagacggggaacgcagtagacgcaaccgacccaaagccca    |                         |
|                |       | ggctcgtcccatatgacgaagcgcctggaatgcatgattaaatccacg    |                         |
| gi 4505999 ref | 210   | AYTGFSSNSERGTAGQVGEPIPTGEAGPSCSSASDKLPRVAKSKFFE     |                         |
|                |       | AYTGFSSNSERGTAGQVGEPIPTGEAGPSCSSASDKLPRVAKSKFFE     |                         |
| genomic        | 26861 | AYTGFSSNSERGTAGQVGEPIPTGEAGPSCSSASDKLPRVAKSKFFE     |                         |
|                |       | gtagtttatgaggggccccgcagggggcttttgtgacccggatattg     |                         |
|                |       | cacgtccacaggcacgatgacgtccgacgcccgaatcgatcacatta     |                         |
|                |       | ccactcccgatgtgacattgtctcttgtgtcctactcggtattgcgctg   |                         |
| gi 4505999 ref | 259   | DSEDESDEAEEDSE                                      | ECSEEDGY                |

|                |       |   |          |
|----------------|-------|---|----------|
|                |       | DSEDESDEAEEDSE  | ECSEEDGY |
|                |       | DSEDESDEAEEDSE  | ECSEEDGY |
| genomic        | 27008 | gaggggtggggggggagGTAAGGG Intron 4 CAGgtaggggt<br>agaaacaacaaaaaga<0-----[27059:27638]-0>aggaaaaga<br>ctgtgatgggaagactg accgagtcc                            |          |
| gi 4505999 ref | 285   | SSEEAENEDEDDETEAEEDDEEEEEEEMVPGMEGKEE<br>SSEEAENEDEDDETEAEEDDEEEEEEEMVPGMEGKEE<br>SSEEAENEDEDDETEAEEDDEEEEEEEMVPGMEGKEE                                     |          |
| genomic        | 27666 | aaggggaggggggagggggggggggaagcgaggagg<br>ggaacaaaaaaacaacaaaaaaatttcgtagaaa<br>ctggagtgatgtccggtagctaaaaaggggaggacagg  |          |
| gi 4505999 ref | 323   | PGSDSGTTAVVALIRGKQLIVANAGD<br>PGSDSGTTAVVALIRGKQLIVANAGD<br>PGSDSGTTAVVALIRGKQLIVANAGD  |          |
| genomic        | 27780 | GTGTGTG Intron 5 AAGcgtgagaaggggcacgactaggagg<br><0-----[27780:28130]-0>cgcaggcccttcttgaatttcacga<br>tctcttaagggcgaaggggtaccaac                             |          |
| gi 4505999 ref | 349   | SRCVVSEAGKALDMSYDHKPEDEVELARIKNAGGKVTMDGRVNGGLNLS<br>SRCVVSEAGKALDMSYDHKPEDEVELARIKNAGGKVTMDGRVNGGLNLS<br>SRCVVSEAGKALDMSYDHKPEDEVELARIKNAGGKVTMDGRVNGGLNLS |          |
| genomic        | 28209 | tctggtgggagtgtattgcacggggcgcaaagggagaaggcgaggcact<br>cggttcacgactatcaaaacaaatatcgtaacggatctaggtaggtatc<br>tctgatgtcatagcttcaagtaaaaaccgtttcgccgtgaccgcccc   |          |
| gi 4505999 ref | 398   | RAI DHFYKRKNLNPPEEQMISAL<br>RAI DHFYKRKNLNPPEEQMISAL<br>RAI G:G[ggg] DHFYKRKNLNPPEEQMISAL   |          |
| genomic        | 28356 | agaGGTAAGGG Intron 6 CAGGGgcttaaaaaccggaatgc<br>gct <1-----[28366:29164]-1> aataagaaatccaaattcct<br>act ccctgacgcgatgagggtact                               |          |
| gi 4505999 ref | 422   | PDIKVLTLTDDHEFMVIACDGI N<br>PDIKVLTLTDDHEFMVIACDGI N<br>PDIKVLTLTDDHEFMVIACDGI W:W[tgg] N   |          |
| genomic        | 29227 | cgaagcacaggcgtagagtggatGGTGAGCA Intron 7 CAGGa<br>catattctcaaaattttcgagt <2-----[29295:29541]-2> a<br>tccgggtctcctacgctcttcc t                              |          |
| gi 4505999 ref | 446   | VMSSQEVVDFIQSKISQRDENGELRLLSSIVEEL<br>VMSSQEVVDFIQSKISQRDENGELRLLSSIVEE+<br>VMSSQEVVDFIQSKISQRDENGELRLLSSIVEEV  |          |
| genomic        | 29546 | gaaacggggtactaaaccggaggccttttagggg<br>ttggaattattacatgagaaagatgttccttaat<br>ggccgatattctaagccgttatggtgagactgagg   |          |
| cDNA           | 1     | MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMEDAHNCIPEL<br>MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSME<br>MGAYLSQPNTVKCSGDGVGAPRLPLPYGFSAMQGWRVSMVRRQGPiGW           |          |
| genomic        | 2134  | aggtctccaagattggggggccccctgttgacgtcgtaggaacgcagt<br>tgcatacactagcgagtgccgtctcagtcctagggtctatggagctgg<br>gtccccgcggggcgccgcggggcccccgacgccccgggggggcacg      |          |
| cDNA           | 50    | DSETAMFSVYDGHGG EEVALYCA<br>++ + G EEVALYCA<br>PLRGGNLTEKESGGW -:G[ggg] EEVALYCA  |          |
| genomic        | 2281  | cccggacagagaggtGGTCCTCC Intron 1 TAGGGgggggtttg<br>ctgggatcaaagggg <1-----[2327:25451]-1> aatctagc<br>gggcgtgggagcgag gatcgctc                              |          |

FIGURE 3, page 16 of 17

|         |       |   |                         |                              |
|---------|-------|---|-------------------------|------------------------------|
| cDNA    | 73    | KYLPDI IKDQKAYKEGKLQK                       |                         | ALEDAF                       |
|         |       | KYLPDI IKDQKAYKEGKLQK                       |                         | ALEDAF                       |
|         |       | KYLPDI IKDQKAYKEGKLQK                       |                         | ALEDAF                       |
| genomic | 25478 | atccgaaagcagtaggaccaGTCTGTC                 | Intron 2                | CAGgtgggt                    |
|         |       | aatcattaaaacaaagataa<0-----[25538:25851]-0> |                         | ctaact                       |
|         |       | attttccatggccgacgagg                        |                         | taatcc                       |
| cDNA    | 99    | LAI DAKLTTEEVIKELAQIAGRPTED EDEKEKVADEDD    |                         |                              |
|         |       | LAI DAKLTTEEVIKELAQIAGRPTED EDEKEKVADEDD    |                         |                              |
|         |       | LAI DAKLTTEEVIKELAQIAGRPTED EDEKEKVADEDD    |                         |                              |
| genomic | 25870 | tgaggataaggggaagcgcaggccagggggagagggggg     |                         |                              |
|         |       | tctacatccaattaatcatcggccaaaaaaatcaaaa       |                         |                              |
|         |       | gttccagctaactaggagtagactgtataaaaattatt      |                         |                              |
| cDNA    | 137   |   |                         | DH FYKRNKNLPPEEQMISALPDI     |
|         |       |   |                         | DH FYKRNKNLPPEEQMISALPDI     |
|         |       | G:G[ggg]                                    |                         | DH FYKRNKNLPPEEQMISALPDI     |
| genomic | 25984 | GGTGAGTG                                    | Intron 3                | CAGGGgcttaaaaacccggcaatgccga |
|         |       | <1-----[25985:29164]-1>                     |                         | aataagaaatccaaattcctcat      |
|         |       |   |                         | ccctgacgcgatgaggtacttcc      |
| cDNA    | 161   | KVLTLTDDHEFMVIACDGI                         |                         | NVMS                         |
|         |       | KVLTLTDDHEFMVIACDGI                         |                         | NVMS                         |
|         |       | KVLTLTDDHEFMVIACDGI                         |                         | NVMS                         |
| genomic | 29236 | agcacaggcgtagagtggg                         | TGGTGAGCA               | Intron 4                     |
|         |       | attctcaaaattttcgagt                         | <2-----[29295:29541]-2> | attg                         |
|         |       | gggtctcctacgctcttcc                         |                         | tggc                         |
| cDNA    | 185   | SQEVVDFIQSKISQ RDENGELRLLSSIVEEL            |                         |                              |
|         |       | SQEVVDFIQSKISQ RDENGELRLLSSIVEE+            |                         |                              |
|         |       | SQEVVDFIQSKISQ RDENGELRLLSSIVEEV            |                         |                              |
| genomic | 29555 | acgggggtactaaaccggaggccttttagggg            |                         |                              |
|         |       | gaattattacatgagaaagatgttccttaat             |                         |                              |
|         |       | cgatatctaagccgttatggtgagactgagg             |                         |                              |